The Impact of Declining Student Persistence in Distance Learning on American College Completion Goals

Francine Adams
Nova Southeastern University, adamsfrancine@gmail.com

This document is a product of extensive research conducted at the Nova Southeastern University Abraham S. Fischler College of Education. For more information on research and degree programs at the NSU Abraham S. Fischler College of Education, please click here.

Follow this and additional works at: https://nsuworks.nova.edu/fse_etd

Part of the Education Commons

Share Feedback About This Item

NSUWorks Citation

This Dissertation is brought to you by the Abraham S. Fischler College of Education at NSUWorks. It has been accepted for inclusion in Fischler College of Education: Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
The Impact of Declining Student Persistence in Distance Learning on American College Completion Goals

by
Francine Adams, EdD

An Applied Dissertation Submitted to the Abraham S. Fischler College of Education in Partial Fulfillment of the Requirements for the Degree of Doctor of Education

Nova Southeastern University
© 2017
Approval Page

This applied dissertation was submitted by Francine Adams under the direction of the persons listed below. It was submitted to the Abraham S. Fischler College of Education and approved in partial fulfillment of the requirements for the degree of Doctor of Education at Nova Southeastern University.

Roberta Silfen, EdD
Committee Chair

Charles A. Schlosser, PhD
Committee Member

Kimberly Durham, PsyD
Interim Dean
Statement of Original Work

I declare the following:

I have read the Code of Student Conduct and Academic Responsibility as described in the *Student Handbook* of Nova Southeastern University. This applied dissertation represents my original work, except where I have acknowledged the ideas, words, or material of other authors.

Where another author’s ideas have been presented in this applied dissertation, I have acknowledged the author’s ideas by citing them in the required style.

Where another author’s words have been presented in this applied dissertation, I have acknowledged the author’s words by using appropriate quotation devices and citations in the required style.

I have obtained permission from the author or publisher—in accordance with the required guidelines—to include any copyrighted material (e.g., tables, figures, survey instruments, large portions of text) in this applied dissertation manuscript.

Francine Adams
Name

August 12, 2017
Date
Acknowledgments

This work would not have been possible without the support of my Dissertation Committee. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. I wish to thank Dr. Charles A. Schlosser for his meticulous demonstration of perfection, which has left a permanent impression on me. I would especially like to thank Dr. Roberta Silfen, the chairman of my committee. As my life teacher and mentor, she has taught me more than I could ever give her credit for here. She has shown me, by her example, what a good scientist (and person) should be. During this process, Dr. Silfen became a friend, oftentimes an only friend who lent me support through many life-changing circumstances.

No one has been more important in the pursuit of this project than my parents Frank Adams and Naomi Adams who instilled in me the desire to achieve beyond my perceived limitations. I am indebted to my brother, Ronald Adams for his diligence and patience in providing professional data architecture reviews. I genuinely thank Darius Tantanella, Erwin Sales Executive, who was key in providing an extended license for this study. I thank Dr. Neil Turner, for being a supportive mentor in this journey, providing me with the benefit of his academic experience and advice. I would like to thank my four children Rhonda Renee Kane, Curtis Warren Adams (aka. Zeus) who never doubted me a single moment, Santeno Kahlil Adams, and Stormy Kim Ramsey. To my grandchildren, Rayana Kane, Dayana Kane, Ceyana Kane, Sanai Kennedy Adams, Jeffrey Curtis Brinkley, and Santeno Lee Adams who did what SugarBabys do; just be you, during those times I needed to connect with family.
Abstract

The Impact of Declining Student Persistence in Distance Learning on American College Completion Goals. Francine Adams, 2017: Applied Dissertation, Nova Southeastern University, Abraham S. Fischler College of Education. Keywords: student retention, student attrition, student persistence, college completion, graduation rate, degree attainment

Academic institution leaders, education researchers, and policy makers have come together to restore America’s ranking among the top nations who are educating citizens to prepare for the technology jobs of the future. President Barack Obama tasked the nation to do so after alarming figures in student achievement were revealed from the 2009 PISA reports. The White House and Lumina Foundation called for a transformation, setting deadlines for 2020 and 2025, respectively, that 60% of Americans should attain degrees or credentials. This study centered on degree attainment by, assessing distance learning persistence for warning signs of impeding these goals.

This study had two components in determining how a collection of disparate data could be assembled for further analysis: one, to collect sample graduation statistics from available primary education data sources. Secondary data sources were reviewed for relevance; and stored for potential future value. The second component was to review a large set of best-evidence research studies (100-200) and identify sample data for identifying or calculating effect size in best-evidence research reports on distance learning persistence to graduation. The two sample result sets were combined in a discovery process of synthesis with data education statistics. Wolf (1986) defined analyses: primary analysis is the original analysis; secondary analysis is revisiting or re-analysis of data to uncover “better statistical techniques or answering new questions”; meta-analysis is the analysis of analyses. This research investigated ways and means to synthesize disparate datasets, utilizing both new and existing data, to garner and/or record aggregate gaps for analyzing student attrition over the term of the goals.

A knowledge base, deemed Best Evidence Statistics Synthesis (BESS1.0) was developed for long term research to capture the effort invested into meta collection studies and for sharing among research teams engaged in those efforts.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Background and Justification</td>
<td>2</td>
</tr>
<tr>
<td>The Research Problem</td>
<td>3</td>
</tr>
<tr>
<td>Deficiencies in the Evidence</td>
<td>6</td>
</tr>
<tr>
<td>Audience</td>
<td>6</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>7</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>8</td>
</tr>
<tr>
<td>Summary</td>
<td>8</td>
</tr>
<tr>
<td>Chapter 2: Literature Review</td>
<td>10</td>
</tr>
<tr>
<td>Introduction</td>
<td>10</td>
</tr>
<tr>
<td>Methodologies</td>
<td>13</td>
</tr>
<tr>
<td>Other Considerations</td>
<td>14</td>
</tr>
<tr>
<td>Sample of Studies</td>
<td>21</td>
</tr>
<tr>
<td>Graduation Rates by Institution</td>
<td>22</td>
</tr>
<tr>
<td>Qualitative Data</td>
<td>23</td>
</tr>
<tr>
<td>Classification of Causes and Solutions</td>
<td>25</td>
</tr>
<tr>
<td>Research Questions</td>
<td>30</td>
</tr>
<tr>
<td>Summary</td>
<td>30</td>
</tr>
<tr>
<td>Chapter 3: Methodology</td>
<td>32</td>
</tr>
<tr>
<td>Introduction</td>
<td>32</td>
</tr>
<tr>
<td>Theoretical Framework</td>
<td>32</td>
</tr>
<tr>
<td>Procedures</td>
<td>38</td>
</tr>
<tr>
<td>Design</td>
<td>39</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>41</td>
</tr>
<tr>
<td>Limitations</td>
<td>42</td>
</tr>
<tr>
<td>Summary</td>
<td>44</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
<td>46</td>
</tr>
<tr>
<td>Phase I: Evidence-Based Research</td>
<td>47</td>
</tr>
<tr>
<td>Phase II: Education Databases</td>
<td>53</td>
</tr>
<tr>
<td>Phase III: Merging the Data</td>
<td>61</td>
</tr>
<tr>
<td>Summary</td>
<td>66</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td>69</td>
</tr>
<tr>
<td>Summary of Findings</td>
<td>69</td>
</tr>
<tr>
<td>Interpretation of Findings</td>
<td>72</td>
</tr>
<tr>
<td>Limitations</td>
<td>73</td>
</tr>
<tr>
<td>Implication of Findings</td>
<td>73</td>
</tr>
<tr>
<td>Recommendations for Further Research</td>
<td>74</td>
</tr>
<tr>
<td>Summary and Conclusion</td>
<td>75</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

For over a decade, the U.S Department of Education posited the country was upholding its global rank among the top 10 nations in reading, science, and mathematics statistics, ever since 1997. When the Organization for Economic Cooperation and Development (OECD) released the (PISA) 2009 Results, the U.S lagged; an alarming 27th on the global rank.


The collective goals of the White House and the Lumina Foundation were for 60% of American citizens to attain quality degrees or credentials by the end of the decade or 2020 and 2025, respectively. President Barack Obama stated that achieving these goals would restore the U.S. as the leader in technological preparedness (Kanter et al, 2011). The White House established state-by-state goals establishing a range of 25-34 years of age, with a deadline of 2020. The Lumina Foundation’s range consisted of students 25-64 year of age, with a deadline of 2025 (The White House, 2011). This study merged those age ranges between 25-64 years of age to encompass a more far-reaching spectrum for analysis. Hereafter, the above-mentioned goals, White House 2020 College Completion Challenge (The White House, 2015) and Lumina Foundation: Goal 2025 (2013) will be referred to as Goal 2020/2025.
The National Student Clearinghouse Research (Fain, 2014) reported that for 96% of college students nationwide the national “persistence” rate was down 1.2% since 2009. “More students are leaving higher education after their first year, according to new national numbers that are bad news for the college completion push” (Fain, p. 1). College completion rates are now high on the national agenda, as is greater complexity in “postsecondary pathways”; there is urgency for new measures of student success outcomes (Shapiro et al., 2015).

This study considered declining student persistence, particularly in distance learning programs and ascertained whether attrition was significantly impacting the attainment of Goal 2020/2025. This study investigated distance education programs for declining student persistence to determine if there was confounding of college completion rates. This study examined the distance education sector to determine whether declining student persistence impacts Goal 2020/2025. The research sought to devise a method to respond to these concerns.

**Background and Justification**

The 2009 OECD-PISA statistics were the initial catalyst for Goal 2020/2025. Three years later, the PISA 2012 report was released, generating further alarm. This prompted the Lumina Foundation (2013) to intensify its strategy for college completion; it would require a more aggressive approach if the nation is to attain the targeted deadlines.

On its Goal 2025 website the Lumina Foundation (n.d.) reported America as 11th in global postsecondary education attainment. The Lumina Foundation reported the percentage of Americans between ages 25 and 64 who earned a two- or 4-year college
degree was 38.7%. Between 2008 and 2010 they reported that rates increased slowly but steadily—from 37.9% in 2008, to 38.1% in 2009, to 38.3% in 2010 (2013, p. 5). By 2015, the Lumina Foundation series, A Stronger Nation through Higher Education, stepped up the strategic pace, concluding; at the current tempo, the U.S. will reach roughly 39 million two-year and 4-year college degrees by 2025. To realize a rate of 60% by 2025, the nation must produce 62 million “high-quality” postsecondary credentials (2015). Lumina Foundation’s (2013) set a new metric for “high quality” attainment. The new metric counts the percentage of working-age (25-64) U.S. adults holding a two- or 4-year college degree or certificate and requires that degrees or credentials are to represent valuable skills and knowledge, as well as the student having to demonstrate the ability to apply them (2013).

The American Institutes for Research Education Policy Center (2013) strengthened its triennial report Education Pays requiring more stringent measures going forward than the traditional national averages. Further, the Council for Adult and Experiential Learning (CAEL, n.d., para. 2) noted that competency-based education suggests, instead of evaluating student progress using the credit hour, students should receive college credit based on their actual demonstration of skills learned.

**The Research Problem**

The problem under scrutiny was potential “aggregation gaps” (Horowitz, 2014) in the metrics for Goal 2020/2025 related to declining student persistence which, if not measured, could potentially confound college completion percentages. Measuring declining student persistence in distance learning programs using standard statistical methodology, such as P-value, was not considered sufficient to afford the specific
conclusions to ascertain whether there is confounding, and whether attainment could be higher than measures indicate. And/or there could be other counts not considered.

Brown, Keppel, Hughes, Hard, and Smith (2013) determined there are gaps in participation and success in distance learning programs. Brown et al. (2013) proposed that those gaps in participation and success are detrimental to both individual lives and to society as a whole.

Reputable sources for education statistics include the International Archive of Education Data (IAED), the National Education Association (NEA), and the National Center for Education Statistics (NCES), the National Student Clearinghouse Research Center (NCS), the U.S. Department of Education (DOE), and the Integrated Postsecondary Education Data System (IPEDS). These sources are abundant with data covering the relevant period (2005-2015) to be considered in this research. The research studies were categorized by type in order to attempt the synthesis with other inputs to this study.

In order to accomplish the goal to measure declining student persistence in distance learning, statistical methods from the above sources were examined for potential “aggregation gaps” (Horowitz, 2014) within some data points specific to this study. For instance, Cook and Pullaro (2010) of American Council of Education (ACE) noted there are inconsistencies in the 1995 6-year graduation rates between the National Center for Education Statistics (NCES) and the Integrated Postsecondary Education Data System (IPEDS). Differences in methodologies for calculating graduation rates are the reason for the discrepancies. This study examined whether discrepancies might persist over time in gathering graduation statistics. According to Sabisky (n.d.) the use of the term
"significant" may be over-rated, in that it does not tell us whether a hypothesis is important in the real world. Further, large databases may not be providing an accurate representation of the figures.

Furthermore, the National Student Clearinghouse Research Center (NCS) (2015) recently invited institutions to include 13 new, yet, optional data elements, to provide more robust reporting. The optionality could result in additional disparity in reporting student success outcomes. If different institutions omit new elements and other institutions include new elements, both do so in varying quantities. It was probable the optional data may not elicit the intended outcome of improving consistency. It was probable potential gaps in the statistics might exclude data useful in aligning with new metrics; or confound percentages by overlooking some declining student persistence when solely examining statistical methods, such as P-value. Qualitative studies might also have proved useful; however, studies conducted employing qualitative methods in quantifying distance learning attrition were scarce, and excluded from this study.

Kanter et al. (2011) conducted evidence-based research to determine effect size for online learning vs. face-to-face learning. The purpose of the study was to draw conclusions on distance education programs for K-12 students. The evidence-based studies are considered a more robust measure for analyzing expert quantitative summations vs. relying solely on P-values. The research did not present sufficient data to form conclusions on K-12 students in distance education. At the completion of the meta-analysis, Kanter et al. concluded online learning appears to be as effective as conventional classroom instruction, but not more so. Their findings, incidentally, were representative of higher education online learning, rather than K-12.
Deficiencies in the Evidence

Evidence-based research on student persistence offered more substantive reasons for declining student persistence, as well as preserving and upholding Goal 2020/2025, which became of grave national importance, as stated earlier by Fain (2014). Considering the recent, but uncertain, H1B Executive Order signed by the White House, potentially limiting foreign workers in the U.S., Goal 2020/2025 could surge in magnitude as a side effect. In a recent review of the literature for online student persistence Hart (2012) acknowledged difficulty in synthesizing the literature. This difficulty frames the use of consistent terminology, where terms like persistence, attrition, and success converge (Hart, 2012). A reconciliation of terms continues in studies on student persistence.

Further, P-value, traditionally, informs the reader whether an effect exists, yet, does not reliably divulge effect size (Sullivan & Fein, 2012). This conclusion validates the preference for pursuing evidence-based research, in addition to identified issues with confounding in past research.

Audience

This study will be meaningful to academic administrators and educators partnering to attain the White House 2020 College Completion Challenge (White House, 2015) and the Lumina Foundation Goal 2025 (Lumina Foundation, 2015) goals, particularly academic institutions with distance learning programs. Academic administrators may be able to better assess improvements in student success rates on persistence in distance learning programs. Statisticians and data scientists may find this study meaningful to the field of data analytics in graduation rates.
**Definition of Terms**

Some of the terms below are known to have alternative, or interchangeable semantics. The definitions were selected from respected sources in the body of research.

**Absolute effect size.** The absolute effect size is the difference between the average and mean outcomes in two different intervention groups. Thus, effect size can refer to the raw difference between group means, or absolute effect size (Sullivan & Fein, 2012).

**Aggregation Gaps.** Data may contain gaps in accuracy due to reporting aggregation. (Horowitz, 2014)

**Attrition.** “The rate at which students terminate college without receiving a degree.” (Tinto, 2013, p. 128)

**Evidence-Based.** Refers to any concept or strategy that is derived from or informed by objective evidence; most commonly, educational research or metrics of school, teacher, and student performance. Among the most common applications are evidence-based decisions, evidence-based school improvement, and evidence-based instruction. (The Glossary of Education Reform (2013)

**Effect-size.** The size of the difference between groups; the main finding of a quantitative study (substantive significance); calculated effect sizes can also quantitatively compare results from different studies and thus are commonly used in meta-analyses (Sullivan & Fein, 2012).

**Persistence.** “Student progress from the student perspective; whether the student continues in higher education” (Tinto, 2013, p. 127).
**P value.** “Informs the reader whether an effect exists (via statistical significance), the P value will not reveal the size of the effect” (Sullivan & Fein, 2012).

**Retention.** “Student progress from the institution perspective: whether the student is progressing through the institution” (Tinto, 2013, p. 127).

**Purpose of the Study**

The purpose of this study was to collect, code, and rank evidence-based studies on declining student persistence to determine whether declining student persistence is a threat to the attainment of Goal 2020/2025. This was accomplished by reviewing evidence-based narratives and recording the purpose, the problem, the research questions and the conclusions, in the first pass. From the pool of evidence-based studies accumulated, those most consequential to declining student persistence in distance education were stored for coding in a second pass. The remaining studies were archived in a table in the database, earmarked in an alternate storage as relevant to the topic of student persistence but not suited for this study. This study will attempt to understand whether declining student persistence is a threat to the attainment of Goal 2020/2025.

**Summary**

America, recognized from the PISA reports (2009, 2011), needed to keep up with the global community in preparing U.S. citizens for the technological future. A mindfulness to address these educational shortcomings became the impetus for setting new goals and undertaking innovative approaches to accelerate the attainment of college degrees and/or certifications. Goal 2020/2025 are the pinnacle of those initiatives and the expectation is these goals will restore America to the top tier in college completion. This study consisted of research in the distance education sector, examining declining student
persistence in support of attaining Goal 2020/2025 by reviewing best-evidence research. The literature review in Chapter 2 furthers this examination.
Chapter 2: Literature Review

Introduction

This literature review investigated the underlying basis for, and metrics of, declining student persistence, college completion rates, and their relationships, if any, to Goal 2020/2025. As mentioned earlier, Goal 2020/2025 are the White House and Lumina Foundation’s (2013) lofty goals to restore America to the top tier in college completion rates for U.S. citizens. Findings on college completion rates suggest there may be potential for “aggregation gaps” (Horowitz, 2014) in statistical analysis. Horowitz defines aggregation gaps as data which may be due to summarizing the detail for reporting aggregation.

Aggregation is often used by organizations to summarize results for brevity, save time, and facilitate discussion. Horowitz (2014) stated aggregation is best suited for identifying patterns on a high level. Public reporting is incapable of reporting everything there is to know about a school or its students (Horowitz, 2014). Data privacy is a factor in the decision to summarize education data, as well. In practice, aggregation does not always require detail to be useful. This research considered whether reconciliation was needed if “aggregation gaps” exist in college completion rates and whether it was possible to dissever aggregated data without compromising legalities such as student privacy.

The decision to undertake this study was made to investigate discrepancies considered as typical, due to disparities in reporting methodologies. This literature review examined those disparities to discern, how and why we capture, analyze and disseminate data in the current methodologies with known disparities. The investigation also
examined whether there were statistical inconsistencies which might be synergized for improved accuracy in reporting of graduation statistics.

This literature review preceded a deeper dive into broad evidenced-based studies investigated, stored, analyzed, and quantitatively coded in this research. Means, Toyoma, Murphy, Bakia, & Jones (2010) conducted research for the U.S. Department of Education, employing the method of by reviewing the abstracts for researching evidence-based studies. A similar approach was used in this study. The studies and databases used as sources were recorded along with their respective statistical procedures and recommended research methods, as the foundation for this research. Each large database was accompanied with pointers to the source for its training to make this tool useful to researchers and for a centralized set of studies specific to an area of education research.

Discovery, from a data perspective, was envisaged to render unspecified, undetermined, or known issues contributing to research anomalies. The investigation and documentation for this study of known anomalies was to be a valuable contribution to reduce some of the time and effort invested in best-evidence research. Issues observed in meta-analytic studies have influenced the decision to opt out of the meta-analysis approach. Rothstein, Sutton, & Borenstein, (2005) warn of potential consequences of publication bias and outcome reporting bias (ORB), respectively, in meta-analysis. Upon first hand review Rothstein et al., defined publication bias in meta-analysis as that which occurs in published literature when a subset of available research differs with all of research in that area. The ramifications, say Rothstein et al., may become a threat to the proponents of the respective body of research.
As an alternative, best-evidence synthesis is the preferred research methodology in education science. The U.S. Department of Education invested in a What Works Clearinghouse (ies.ed.gov) in 2002, where scientific evidence based studies are captured for resource and review. The What Works Clearinghouse serves as a model for the fundamental vision to build upon.

Upon drilling into the recesses of existing research, studies were interpreted for relevance to distance learning persistence, coded and ordered by clustering and ranking. Positive and negative rankings were assigned and applied to evidence-based outcomes. Further, studies were categorized by content and contribution to this research. Hart (2012) compiled a literature review meaningful for grasping the range of terminology (both synonyms and antonyms) to define student persistence. The variations were such that, in some literature, persistence may be the opposite of attrition. In other cases, they may be synonymous, in which case some reconciliation was required. Terms used synonymously (attrition, persistence, retention) differ by author perspectives; e.g., institution vs. student perspective (Hart, 2012). The dilemma was that not all researchers concur on these terms throughout various studies. It was the hope of this researcher that a pattern of convergence would emerge in the collection of evidence-based studies.

Accordingly, a mixed method investigation into student persistence and support services in distance education by Brown et al. (2013, citing Boyle, Kwon, Ross & Simpson, 2008) reported there were lower institutional retention rates in online learning than conventional face-to-face education. In their intensive evidence-based study on distance education programs, Means et al. (2010) gathered studies with random-assignment or controlled quasi-experimental designs which examined effects of objective
student learning. Means et al. (2010) concluded online learning to be as effective as traditional classroom instruction; however, they did not find it to be more effective.

**Methodologies**

Means et al. (2010) adopted the view of modern medicine (Sullivan & Feinn, 2012) that evidence-based research, using effect size, affords more exacting conclusions than P-value alone. Increasingly, across various disciplines, P-value continues to decline as the preferred methodology. Perhaps researchers engaged in future studies may re-evaluate where and how to assess the body of education research from past studies, which have relied more heavily upon statistical methods and using P-value? This research might indirectly address that question. This study recorded and ranked education database statistics with the intention to integrate with best-evidence research.

Means et al. (2010) collected 1132 abstracts which were whittled down to 176 studies meeting their criteria of experimental or quasi-experimental design, with objective student learning outcomes in distance learning. Their studies consisted of higher education research in distance learning despite the research team’s intent to report on K-12; however, there were insufficient data.

Sullivan and Feinn (2012) continued with effect size and absolute effect size analysis and considered transformed measures to be a friendlier scale (than P-value); P-value will only report whether an effect exists. Evidence-based research (Sullivan & Feinn) and “aggregation gap” analysis (Horowitz, 2014) were amalgamated, to test this hypothesis.
Other Considerations

The Lumina Foundation (2015) examined prolonged, yet lingering, increases in college completion rates between 2008 and 2011. They concluded the increases were not meeting the anticipated U.S. levels for postsecondary attainment. By their 2015 assessment the Lumina Foundation determined the numbers at the current pace were inadequate to reach projected goals for degree attainment. For the nation to attain Goal 2020/2025 college completion rates must accelerate well beyond the Lumina Foundation 2015 measures. These measures must map to new metrics, if applicable, and accurately inform of failure or success by global measures. Going forward, measures should, ideally, uncover potential “aggregation gaps” (Horowitz, 2014) and potentially discover new niches not considered during goal setting.

The U.S. ranked first in the world, in 1990, for 4-year college degrees. The 2009 PISA reports roused the U.S. Department of Education Secretary and the White House to act, academic institutions responded to the challenges. Despite the aggressive goals for degree attainment, Cook and Pullaro (2010) found numerous discrepancies in college graduation statistics. Many of the databases reporting on college completion rates provide statistics for a 6-year span. The 6-year standard may not be adequate to accurately capture U.S. college completion rates. Hess, Schneider, Carey, and Kelly (2009) researched college completion rates by academic institution. They were unable to validate their results due to the known discrepancies between sources. Most of those databases do not consider those students who persist beyond 6 years when measuring college completion (Cook & Pullaro, 2010).
Due to the OECD-PISA reports (2010, 2012), the White House has challenged Americans to complete at least 1 year of higher education, setting goals by state. The challenge was to reconcile those statistics to the reports which most accurately reflect the combined efforts of the Nation in a fashion that is unabridged.

Preliminary analysis, coupled with prior discovery, suggested synthesis of data in an undertaking of this magnitude, is no simple feat; perhaps, an impossible feat. Yet, it still to begged the question, why do analysts, policy makers, institutions, and potential students rely on “significantly inaccurate” data (Cook & Hartle, 2011) for critical decision making. Data aggregation is a convenient IT systems method to conserve space and cost (Marvasti, n.d.). Nevertheless, there is a cost associated, since Marvasti (n.d.) concluded:

Specifically, 50 percent of IT type metrics lose their distribution coherence (relative to the raw data) after 2.5 hours of aggregation and 85 percent lose their distribution coherence after 12 hours of aggregation. Thus, for capacity studies that rely upon a faithful representation of the raw data distribution, any aggregation level above 2.5 hours may severely impact the final results. (p. 11)

Marvasti (n.d.), said researchers, analysts, policy makers and others who use an organization’s data assets, typically expect to suffer loss in data accuracy, as a matter of course, in the transformation from raw to aggregated data. In current data analysis, this is accepted. He identified several quantification techniques suggested for information loss attributed to data aggregation. The techniques outlined by Marvasti (n.d.) may, or may not, be applicable in this study; however, his work validates that there is reasonable
justification to consider “aggregation gaps.” Hess et al. (2009), researching by academic
institution (see below), acknowledged the findings of outliers. These were cases of
students who were outside of the 6-year reporting standard for 4-year institutions still
struggling to earn enough credits to graduate. Some of those students dropped out of
college altogether, or transferred. According to the way graduation rates are counted, data
for students who dropped out were no longer visible. Clark and Avery (2010) ascertained
aggregation could be defined as information loss occurring as a side effect of substituting
macro-level data for micro-level data. Information loss is a common statistical issue in
the aggregation methodology usually ignored in favor of results.

How was this methodology arrived at, where aggregation issues are ignored in
education data statistics? In 1996, the National Postsecondary Education Cooperative
(NCES) formed a Unit Record vs. Aggregate Data Working Group to adopt a standard
for reporting. The Working Group was tasked to:

(a) contrast and evaluate benefits and limitations of unit record level reporting
versus aggregate reporting; (b) identify and analyze factors that make unit record
level reporting an issue, including confidentiality, flexibility with changing
requirements and definitions, costs, and burdens; (c) recommend new approaches
for collecting, maintaining, and sharing data taking into account technology,
program delivery, and other changes affecting postsecondary education; (d)
document and evaluate prominent unit record level and aggregate reporting
processes and practices; and (e) document and evaluate record sharing practices at
the unit level, as well as, the aggregate level between institutions, governing
boards, and state and Federal governments. (NCES,1998, p. 1)
Detail data vs. aggregated data translates to the protected status of unit level data, available only in special circumstances. Privacy concerns supersede transparency of the data; therefore, those concerns must be held to accountability as related to this research. Privacy may be the staunchest factor in these scenarios; however, this research sought to uncover strategic points of improvement. According to the National Center for Education Statistics (1998), aggregate data are data collected at the institution level, while unit record level data was collected by student record. The report aided in establishing there are relevant concerns with the issues encountered in relying heavily on aggregated data collection. Among those were:

(a) data initially collected at the institutional level cannot be used for lower levels of aggregation, for example, the tracking of individual students over time and across institutions/activities, and (b) aggregate data systems lack the flexibility to examine relationships among variables and to re-aggregate data, should reporting needs change. (NCES, 1998, p. 3)

The 1998 report represents a “point in time” and reasoning in an era which was not technically advanced, to a degree perhaps inconceivable at that time. In fact, a portion of the report examined paper document collection vs. electronic collection. Aggregation disparities persist, validating the prospect expressed in 1998 that current data collection and retention practices would enable or constrain future uses of the data and their value in these future applications. This study noted recording inaccuracies, and divulged some recommendations for direction encouraging more efficacious analysis.
In 1998, standards were rather archaic by today’s methods, such as, examining paper documents vs. electronic collection. The use of archaic standards may be stifling America’s ability to progress if still embedded in policy and practice. These standards could be unintended contributors to America’s slowing or stagnation in college completion goals. The prospect of a supplementary approach; or sector of an approach, which could serve in part, some future, constructive sum of practice in assessing educational attainment, was abundant reason to pursue this challenge. Most of the education data on graduation statistics disclosed figures for degree attainment from conventional institutions. New statistics have become available exclusively on distance learning at the National Center for Education Statistics (NCES, n.d.) beginning in 2012 through 2015, which bolstered this study with fresh perspectives. There might be online education institutions in existence with students who could contribute to the numbers, yet are not being captured. In this climate of education crises, there are may be more parents engaging in home-schooling and/or K-12 distance learning institutions which could afford additional aggregate gaps for entrants from non-traditional sources not captured in graduation statistics. There may be religious sectors with academic institutions which do not participate in federal statistics incentives.

The assumed confounding may be attributable to the plethora of complex and analytically challenging (Cook & Hartle, 2011) graduation statistics reported from various sources. According to Cook & Hartle, some of the composite uncertainties were deemed significantly inaccurate by Integrated Postsecondary Education Data System (IPEDS), resulting in excluding sets of students from some calculations. Further, the Higher Education Act of 1965 (HEA) (Pub.L. 89–329) was rewritten no less than eight
times. This act, along with the Student Right to Know Act of 1990 (Civic Impulse, 2017), are the standards for reporting graduation statistics. This research also considered that the task of collecting, analysis, and reporting on graduation rates has outgrown the methodology (Cook & Hartle, 2011). This consideration is outside the realm of this study; however, it is recommended for more definitive investigation. Hess et al. (2009) demonstrated research from a different perspective (by institution rather than state); however, imperfections in the data sources impeded the accuracy of outcomes.

Assembling and synthesizing the data from education databases was a challenge of reconciliation, coupled with a synthesis of distance learning persistence findings, coded for statistical analysis. Thought should also be given to exploration begun recently by The Institute for International Education (Belyavina, Li & Bhandari, 2013). The Open Doors exploration, say Belyavina & Bhandari (2013), revealed some 320,000 American postsecondary students pursuing various education experiences abroad. Some 273,000 were American students engaged in supplementary study abroad. Among those numbers were 46,500 United States citizens pursuing a degree abroad in 2011/2012 (Belyavina & Bhandari 2013). The Institute for International Education’s data on U.S. Students in Overseas Degree Programs is significant. This was a first-time endeavor to comprehensively account for United States students pursuing degrees abroad at all levels of postsecondary study (Belyavina, Bhandari, 2012). Further, it was relevant to this study that this student sector has never been fully captured.

The above-mentioned research supplemented the numbers for degree attainment sought by Goal 2020/2025. Further, the inclusion in this study offered more constructive views that greater numbers of U.S. citizens are prepared for the technological future
within the U.S. and abroad, even if dispersed. Degree bearing immigrants who become U.S. citizens during the time frame of Goal 2020/2025 would also contribute to increases in the numbers for attainment.

The numbers of American citizens completing their degrees overseas due to the financial attraction is somewhat vague at present, although it is being tracked by Open Doors. The discovery process produced significant data for storing in the knowledge base. Burkholder & Holland (2014) researched international retention and persistence data and reported on figures for Australia, Scotland, the United Kingdom and other countries, as well as the continent of Africa. International research on retention and persistence, noted Burkholder and Holland, focused on college preparation, finances, and high-quality faculty. From the distance learning perspective, there could still be “aggregate gaps” in the distance learning arena for students attending global institutions, especially online, while remaining U.S. citizens and seeking technical jobs in the American workforce. If there is a sector of college graduates who are not on the grid, this research did not directly answer that question. This research was an opportunity to build upon the question and opportunities for further examination. While America is pursuing college completion goals and researchers have confirmed there are concerning incongruences in reporting graduation statistics, then perhaps a new approach is warranted.

Some studies gathered for this research reference the work of Vincent Tinto, who constructed a model in 1975 on conventional education dropout perspectives from a synthesis of dropout rates. Davidson and Wilson (2013) concurred that Tinto’s framework, which focused on student integration, is insufficient to address the
idiosyncrasies of non-traditional student populations in today’s climate and for this research online learning, as well.

Tinto (2013) provided useful definitions that suggested a set of definitions. Some of Tinto’s definitions were useful in this study (see Definition of Terms, above). In this publication, Tinto (2013) closed the circle on his theory of student retention. Tinto’s (2013) theory was useful in this study on a generic level, such that endorsement of a 3-year cut-off may be insufficient for 2-year colleges, while a 6-year cut-off may not suffice for 4-year colleges to draw conclusions on dropout rates. Hess et al. (2009) concurred in their research on graduation rates by institution. Beyond the generic parameters, the theory was not applicable to this research, such as the physical campus classroom serving as home base to encourage attrition (Tinto, 2013). As for improving graduation rates, Tinto (2008) offered sound principles including “systematic, systemic preparation” and offered a suitable case study at the University of Florida, where rates improved simply by making courses available based on student needs, rather than instructor needs.

**Sample of Studies**

There was a generous pool of studies on distance learning persistence. Many were comparative with face-to-face studies or dealt with attrition by other countries standards and, therefore, were archived. In comparative studies on distance learning vs. conventional learning Means et al. (2010) concluded online learning was as effective; however, has not surpassed conventional learning, their research, unintentionally, focused on higher education. On the other hand, a review of the pool of studies in distance learning persistence conceded the dropout rate was higher in distance education than
conventional learning (Boyle, Kwon, Ross, & Simpson (2010) and mandated solutions. More insightful analysis was performed with the intent to dissect the data and record similarities in the findings.

Jenkins (2012) conducted an audit to validate the 2009 meta-analysis by the U.S. Department of Education on distance learning. Jenkins asserted there were numerous flaws in the methods of analysis that warrant further scrutiny. Jenkins assertions encourage careful inspection in the selection of studies for this or future best-evidence or meta-analysis researchers may wish to undertake.

**Graduation Rates by Institution**

Hess et al. (2009) researched graduation success rates by college institution, rather than by state. Their approach, utilizing Department of Education data from 1400 institutions and Barron’s Profiles of American Colleges, observed the data from a perspective that was useful to this study. The researchers’ approach was innovative and inspiring:

A “competitive” category… consists of students with average scores of 500 to 572 on each of the SAT’s three sections (math, critical reading and writing) and GPAs that range from C to B-; they tend to accept 75% to 85% of their applicants… When the 660 schools in this category were ranked by graduation rate, the average for the bottom 10 places was shockingly low–only 20%. In contrast, the average graduation rate for the top 10 schools in the category was about 75%... Put simply, students with roughly the same credentials were three times more likely to graduate in six years, if they had gone to one set of schools rather than the other. (Hess et al., 2009)
Hess et al. (2009) identified certain imperfections, in what could be described as limitations in the data, rather than the research. The possibility was, much could be gleaned from their research that may serve as a springboard for future reporting, along with justification. Hess et al.’s research on graduation statistics data by institution encouraged a deeper dive, to potentially ascertain whether there can be functions, transformations, and reconciliations to proceed further on their approach.

**Qualitative Data**

Russo-Gleicher (2014) concluded that online college course completion rates, at 10-20%, were significantly lower than traditional courses. The study consisted of interviews with 48 teachers at a college in the Northeast in 2012. It concluded that 50% of the participants were dissatisfied with screening student readiness for online education; in addition, capability (technical) and emotional maturity were significant in student success (Russo-Gleicher, 2014). Online teachers who were interviewed had areas of dissatisfaction to contend with factors that are not found in traditional classrooms, such as access to the Internet and previous online experience. The framework of distance education may continue to have obstacles to success.

Harrell & Bower (2011) conducted a study on factors contributing to student success in online college courses. Their research findings for success were similar to the research areas of dissatisfaction (Russo-Gleicher, 2014), which covered learning style, locus of control, computer experience and access, previous online experience, and various personal attributes. Although there are differing views on student success there were many similarities across studies with elements that were included or excluded in other studies.
Elements of relevance that fell between the studies conducted by Harrell & Bower (2011) and Russo-Gleicher (2014), respectively, suggested there may be justification for further inquiry, and reason to record, and store findings not necessarily specific to this study. Those findings could serve future analysis, which may prove valuable beyond this research. If specific elements to a study are omitted and generic elements of data, relevant to a topic, are preserved in sets, then inspecting a set of studies may be beneficial to advance the progression of searches for a researcher or research group. This is essentially the purpose for the database also considered a knowledge base.

For example, the below findings by Hall (2016) are on high school graduation rates. The source for these data was stored to access information, along with other research on high school data. Even though not utilized in this study, the set of data retained may prove beneficial to future studies. This would be a contribution of ‘sets of research’ accessed, then added to the knowledge base on said topic:

- 33 states graduate less than 70% of their students with disabilities; seven of those states graduate less than 50% of students with disabilities.
- 11 states graduate less than 70% of Hispanic/Latino students.
- 17 states graduate less than 70% of African American students.
- 16 states graduate less than 70% of low-income students. In those states, researchers estimate that nearly 191,000 low-income students did not graduate on time with a regular diploma.
- 35 states graduate less than 70% of English-language learners; seven of those states have ELL graduation rates under 50%.
• 10 states graduate less than 70% of all five subgroups. They are Colorado, Georgia, Michigan, Minnesota, Nevada, New Mexico, New York, Ohio, Oregon, and Washington (Hall, 2016).

A collection or set of high school data could be categorized, stored and retrieved, as needed. A database dedicated to best-evidence statistics synthesis will facilitate a less work-intensive process for future researchers reusing the set of studies in their research domain. It became apparent, in a high-level overview in the meta-analysis approach, that many of the same studies are repeatedly extracted by many researchers, seeking unique outcomes. As an example, Tinto’s work from 1975 forward continues to be referenced in studies on student persistence. In doing so, there exists a set of Tinto works, not all of which were included in this research. It would also be possible to cross reference researchers who have sets of studies on unique topics.

Typically, studies are gathered from various databases for review. Select elements are later extracted for compilation in the respective research. The content collected for research may or may not be stored for various reasons; however, the studies can be catalogued in sets that can be retrieved by researchers, furthering their efforts to gather data. Updated large datasets that are available for download could then be uploaded when new relevant data or statistics become available. Figures such as high school graduation rates (Hall, 2016) could be captured as relevant content and accessed as needed. This database can be constructed in a performance-efficient manner.

**Classification of Causes and Solutions**

Schectman, Debarger, Dornsife, Rossier, & Yarnell (2013) developed a hypothetical model for grit, tenacity and perseverance demonstrating there are
multifaceted concepts for goals, challenges, and ways of managing them. This model has an attraction due to the strength and admirable qualities it advocates which students can adopt for persistence. These qualities were advocated to enhance self-esteem and promote success. Schectman et al. identified optimally challenging goals within every student’s range of proximal development.

Given an inherent model such as Grit, Tenacity and Determination coupled with a success plan to employ these, Schechtman et al. offered students a roadmap to provide themselves with motivation to succeed, thereby supplementing the resources in their institution and the efforts of their instructors. They demonstrated that the best tools to tap to promote student success are inherent at the outset.

Schechtman et al. (2013) endorsed three malleable and teachable facets: academic mindsets, effortful control, and strategies and tactics. The Schechtman et al. model offered a non-academic learning track that any student may adopt and practice without grade anxiety or performance anxiety when owned and voluntary. Pressure to adhere, when driven to achieve a goal (Schechtman et al) could, potentially, result in anxiety. Measures, tools, indicators developed around evidence (Schechtman et al.) would aid in identifying at-risk students who may be better served by an alternate track. There are constructs in learning science with numerous measuring tools capable of tracking nuances to provide more in-depth comprehension (Schechtman et al., 2013). Some of the factors in need of solutions are learning curves, training, privacy, and resource intensity for busy educators (Schechtman et al., 2013).

More attention was devoted to this hypothesis during this study. This research could potentially result in a model for addressing declining student persistence in higher
education in general. Alternately, the knowledge base developed could result in a tool to enhance search methodology. The application in distance education may be worthy of further consideration, as well as replication.

The following is the hypothesis of the body of contextual factors and psychological resources:

Students pursue a wide range of goals and encounter many different types of challenges and setbacks. Socio-cultural context can play an important role in determining what students value and want to accomplish, the types of challenges they face, and the resources to which they have access. Their perseverance may be directly influenced by contextual factors in the learning environment and can require engagement of important psychological resources—academic mindsets, effortful control, and strategies and tactics. (Schechtman et al., p. 16)

This hypothetical model suggests there are socio-cultural contexts for inherent resources students can manifest and cultivate on their own, given the proper motivation. The hypothesis contains inspirational buzzwords, such as self-reporting, motivation, and integrity that would be expected to have positive impacts on students aspiring to succeed.

Schechtman et al. (2013) tackled the potential for a counterproductive outcome when applying the concepts of grit, tenacity, and perseverance. There are certain risks associated with lack of cognizance in educators or parents on how to apply this, in practice, which could cycle the outcomes to return to stress and/or anxiety. This could if educators and/or parents start overrating grit, tenacity and perseverance. Schechtman et al. contrived a hypothetical model to demonstrate the concepts (Figure 1).
Following the investigation into grit, tenacity and perseverance, and the creation of the hypothetical model, Schectman et al. compiled a list of 8 conclusions (Table 1) regarding their findings. Among those were some of the potential issues, recommendations for utilization with new technologies, concerns for overstating, as mentioned above, in utilizing the concepts, and a need for researchers to develop working models to implement grit, tenacity and perseverance to facilitate success in education.
### Table 1

**Grit, Tenacity, Perseverance Conclusions**

<table>
<thead>
<tr>
<th>Conclusions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>For significant and pervasive shifts in educational priorities to promote not only content knowledge, but also the non-cognitive factors of grit, tenacity, and perseverance, there is a strong need for growing involvement and support by all educational stakeholders.</td>
</tr>
<tr>
<td>Two</td>
<td>Substantial research points to actionable “best practices” to promote grit, tenacity, and perseverance. Note that there is still limited evidence at scale, and the field still needs coherent methods for integrating these practices into school culture, teaching practices, curriculum, and technology—especially under conditions that present significant barriers.</td>
</tr>
<tr>
<td>Three</td>
<td>Educators and researchers have demonstrated important successes in promoting grit, tenacity, and perseverance through brief interventions, teacher professional development programs, alternative school models, informal learning models, and digital learning environments. New and emerging technologies can provide opportunities for optimal challenge through adaptability, promote academic mindsets, teach learning strategies, promote the development of effortful control, and provide motivating environments.</td>
</tr>
<tr>
<td>Four</td>
<td>In this accountability-driven culture, there are a wide range of systemic and structural barriers that prevent broad implementation of many best practices and programs. Limitations include short school periods with broad coverage of standards, lack of teacher training, lack of time for teachers to plan and collaborate, and lack of parental support.</td>
</tr>
<tr>
<td>Five</td>
<td>While there is a great deal of work in this area broadly, the importance of grit, tenacity, and perseverance in education is not necessarily widely known, and stakeholders at many levels may not understand the importance of investing resources in these priorities. In many settings, awareness-raising is necessary so that teachers, administrators, parents, and all other stakeholders in the educational community see these issues as important and become invested in supporting change.</td>
</tr>
<tr>
<td>Six</td>
<td>Parents and guardians can also play a direct and important role in promoting their children’s grit, tenacity, and perseverance. A systematic exploration of the complex roles of parents and the home environment was outside the scope of this report. However, some important themes did emerge in our interviews.</td>
</tr>
<tr>
<td>Seven</td>
<td>Across communities of practice and research traditions, inconsistency in conceptual terminology is a barrier to collaboration and progress.</td>
</tr>
</tbody>
</table>
Eight There is a need to develop empirically based models of pathways for developing grit, tenacity, and perseverance over time, in different contexts, and for different types of goals and challenges. Such work would inform the development of learning trajectories and selecting age-appropriate and context-appropriate interventions.


**Research Questions**

The data compiled and synthesized will be analyzed to answer the following questions:

1. How can evidence-based research outcomes be ranked and coded for data analysis to merge and measure, in parallel with traditional statistics and Goal 2020/2025 metrics to determine if there is an impact from declining student persistence outcomes in distance education on college completion rates through 2025?

2. How can statistics for college completion from 2005 to 2015 be disaggregated to compare, in parallel, with evidence-based studies in an analysis to consider declining student persistence outcomes in distance education from 2005 to 2025?

3. How can statistics for college completion from 2005 to 2015 be disaggregated to compare, in parallel, with evidence-based studies in an analysis to consider increases in student persistence outcomes in distance education from 2005 to 2025?

4. How can the metrics for Goal 2020/2025 be quantified to integrate declining student persistence outcomes in distance education to ensure there are
no “aggregation gaps” (Horowitz, 2014) in the percentages for college completion due to declining student persistence between 2005 and 2025?

**Summary**

The literature revealed there could be a set or sets of hypothetically related factors in student persistence, which could impact the attainment of Goal 2020/2025. To best examine the potential impact this research scrutinized areas of declining student persistence, as well as other areas for transparency to determine if significant to this study. The U.S. Department of Education has determined scientific evidence-based studies to be the preferred methodology for education research, this was selected as the approach.

The history of aggregated data vs. detail data was reviewed and it was determined that further investigation is warranted to research the usefulness of dated norms in an era of accelerating technology. There was also a sector of students taking distance education courses abroad which may not be accounted for in graduation statistics. The above factors were deemed important to this research, yet still preceded the research into scientific evidenced based studies gathered to reveal additional findings important to this study.
Chapter 3: Methodology

Introduction

One of the goals of this study was to attempt to meet the caliber of excellence sought in education research which favors scientific best-evidence studies per the U.S. Department of Education. Formerly the domain of medicine and psychology, gathering best-evidence research and compiling statistics for synthesis is gaining increasing acceptance and value in research in general. The preliminary approach was employed in a study conducted by Means et al. This study expanded on their approach by attempting transformations of statistics extracted from the studies collected and progressed to attempt a synthesis with graduation statistics. Select education databases were identified for data relevant to this research.

Theoretical Framework

To determine the best-evidence-based methodology, changing trends in education reform were considered. As per the methodology employed by Means et al. (2010), the abstracts of evidence-based studies were reviewed. Using Means et al.’s methodology, the abstracts were interpreted for relevance. In this study, those abstracts with relevance to student persistence in distance learning were selected, and the highest level of
experiment and objectivity were then ordered by clustering similarities in factors: purpose, methodology, outcomes and statistical references.

The Great Schools Partnership (2013) and the U.S. Department of Education (DOE) place increasing emphasis on excellence in research methodology, particularly in education, by providing evidence-based justification. Great Schools Partnership (2016) asserts:

A widely-used adjective in education, evidence-based refers to any concept or strategy that is derived from or informed by objective evidence—most commonly, educational research or metrics of school, teacher, and student performance. Among the most common applications are evidence-based decisions, evidence-based school improvement, and evidence-based instruction. (2016, para. 1)

This study consisted of a synthesis of evidence-based studies and education research databases to understand how to do a quantitative summation of the effect-size from the combined studies on declining student persistence. Upon discovery and selection of a suitable format, non-conforming data was transformed to attempt consistency for integration, consolidation and analysis. Evidence-based research (Sullivan & Feinn, 2012) and “aggregation gap” analysis (Horowitz, 2014) were examined, in combination, to test this hypothesis. Cook and Pullaro (2010) disclosed discrepancies in college graduation statistics. Many of the databases reporting on college completion rates provide statistics for a span of 6 years. Few researchers report on students who persist beyond 6 years to completion.

There was an in-depth document research. Data were collected in the form of electronic and paper-based documents from evidence-based studies and statistical data
from databases reporting on distance learning. The study consisted of searches for evidence-based studies on distance learning in specific databases, including: ProQuest, Springer, Sage, ERIC, and Google Scholar. Additional sources of academic value were added to this list, such as relevant international reports. Statistical data were gathered from relevant large education research databases.

The study included best-evidence research or education statistics relevant to this study. This included some historical data for to recognize which methods could be useful to a synthesis of best-evidence research with education database statistics.

A set of databases was subjected to a preliminary review. Barron’s Profiles, although not an education database, is of interest due to the innovation of efforts by Hess et al. to analyze student data by institution, although not within scope of this study. The Integrated Postsecondary Education Data System recently began to include statistics on distance learning data which, as mentioned above, was timely for his study. National Student Clearinghouse data were also examined for figures on 2-year and 4-year students.

During this study one novel finding was the innovative survey conducted through the Institute for International Education which collected data on U.S. students who completed degrees oversees in 2009. This survey captured data that have not been collected before. This research examined the aforementioned survey along with students who are U.S. citizens pursuing degrees online who are not based in the U.S. and may also represent students who have fallen through the cracks of graduation statistics (Table 2).

Selected tools were IBM SPSS and R, a plug-in which extends the capability of SPSS, the statistical analysis tool. Microsoft SQL Server Express, a relational database and QlikView, a business intelligence application for analysis were also selected for
testing. CA ERwin Data Modeler was provided by erwin. The erwin salesperson extended the complimentary trial specifically for this research. It was useful, not only to this research, but for future consideration to test the data modeling and reverse-engineering capability in CA ERwin. CA ERwin Data Modeler is recommended due to its ability to reverse engineer a database and port the content to a different database. The many disparate sources in Table 2 are indicative of the complexity that could arise in collecting data.

Table 2

**Education Databases**

<table>
<thead>
<tr>
<th>DATABASE</th>
<th>DATA CONSIDERED</th>
<th>RANGE AVAILABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barron’s Profiles</td>
<td>Profiles of American Colleges data for alternate non-state level views</td>
<td>TBD</td>
</tr>
<tr>
<td>Digest of Education Statistics (NCES) Annual Reports</td>
<td>Digest includes the number of institutions, teachers, enrollments, and graduates; educational attainment, finances, and federal funds for education</td>
<td>2005-2015</td>
</tr>
<tr>
<td>Institute for International Education</td>
<td>U.S. students seeking degree completion abroad</td>
<td>2014-2014</td>
</tr>
<tr>
<td>Integrated Postsecondary Education Data System (IPEDS)</td>
<td>New Data collection by distance education participation</td>
<td>2011-2016</td>
</tr>
<tr>
<td>Integrated Postsecondary Education Data System (IPEDS) Distance Learning Data</td>
<td>Students enrolled in degree-granting postsecondary institutions, by distance education participation</td>
<td>2011-2015</td>
</tr>
<tr>
<td>National Science Foundation (NSF)</td>
<td>The National Survey of College Graduates (NSCG) a longitudinal survey to provide data on the nation's college graduates.</td>
<td>2005 - Feb 2015 (release Sep 2016)</td>
</tr>
<tr>
<td>National Student Clearinghouse Research Center (NCS)</td>
<td>6 YR Start/Completion Rates</td>
<td>2005-2014</td>
</tr>
</tbody>
</table>
Open source software was desired whenever possible, due to the autonomy to customize the software to the problem without concern for licensing violations. To innovate and garner feedback from the open source community would also enable researchers to afford the tools, if there is a lack of funding sources. In retrospect, the open source applications were a nice to have, but the learning curve was not well suited for this study.

A background in data management and over 20 years’ experience in business intelligence were the motivation to develop a knowledge base/research synthesis tool for the research community that is user friendly, yet broad and comprehensive. A fellow data architect served as a consultant for validation, verification, and feedback. A statistician was available for the statistical validation of the experimentation. The selected database was scalable to accommodate future increases in the size of datasets, over time with the least complexity. PostgreSQL was chosen for its robust sophistication as an open source database and simplicity to of its installation. Eclipse: BIRT was considered as the data analysis tool for ease of acquiring and installation. CA ERwin is a moderately priced commercial data modeling tool. The vendor, erwin, graciously extended a trial for this study. CA ERwin is highly recommended as a versatile tool with extended capabilities,
support structure and is the preferred tool in the industry for research of this nature. The preferred statistical tools were IBM SPSS and R, the add-in which extends the capability of IBM SPSS. See Table 3 below.

The software was reviewed and installed on a desktop PC with Windows 7 operating system. A long-term perspective on the capability of the tools is of prime importance in developing a suitable infrastructure. While this study is not focused on the architecture, the infrastructure for the tool is of importance for ease of use and efficiency. Ideally, data storage would be on a dedicated server for security, performance, updates, and to protect and preserve the collaboration and confidentiality while making it easy to add new research team members.

Table 3

**Bess Application Review**

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Cost</th>
<th>License/Support</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIRT</td>
<td>Business Intelligence</td>
<td>Free</td>
<td>Open Source</td>
<td><a href="http://www.eclipse.org/birt/documentation/">http://www.eclipse.org/birt/documentation/</a></td>
</tr>
<tr>
<td>CA ERwin</td>
<td>Data Modeler</td>
<td>$1,259</td>
<td>CA ERwin Support</td>
<td><a href="http://erwin.com/products">http://erwin.com/products</a></td>
</tr>
<tr>
<td>PostgreS</td>
<td>Database</td>
<td>Free</td>
<td>Open Source Community</td>
<td><a href="https://www.getapp.com/business-intelligence-analytics-software/a/pentaho/features/">https://www.getapp.com/business-intelligence-analytics-software/a/pentaho/features/</a></td>
</tr>
<tr>
<td>IBM IPSS</td>
<td>Statistical Tool</td>
<td>$1,170</td>
<td>IBM Support</td>
<td></td>
</tr>
</tbody>
</table>
The analysis would, ideally, include an assessment of the amount of data, the amount of storage, and the capability of the desired tools to withstand processing stress and query analysis. These are all considerations for future enhancement. It is possible there may be bountiful opportunities resulting from capturing information that would be timesaving, laborsaving, and cost-effective in a research knowledge base.

**Procedures**

The procedures for this study were:

1. Data exploration consisting of using search terms to find a substantive number of quantitative studies to conduct a preliminary narrative review of the abstracts of the relevant independent studies. Search terms used were: student persistence (or attrition) in distance learning (or education) and online learning (or education). Studies were initially filtered to collect research conducted between 2010 and 2016.

2. To further isolate the relevant subset for a quantitative, summative synthesis of best-evidence-based narratives, the abstracts from the set were reviewed for inclusion or exclusion for this research. Those included were ranked by selected criteria in a first pass. Those excluded were archived and stored in an alternate folder by unique ID for future accessibility.

3. More definitive criteria were defined to prioritize those quantitative studies which examine student persistence in distance learning. Once identified, the studies were clustered for analysis.
4. Refereed journals and government legislation were reviewed, using the same approach as above, to glean criteria for clustering. Once documents were included from these sources, they were grouped with best-evidence studies rather than with education databases.

5. Statistics were collected from the selected education research databases. The statistics in this research were summary statistics to inspect the data structure for usefulness in BESS1.0. Detailed statistical data, although preferred for deeper inspection for synthesis and gap identification was not useful to the purpose of this study. While discovery may be in the detail, aggregate data were examined in this preliminary approach to identify aggregate gaps.

**Design**

Figure 2 is the graphical depiction of the processes and data flow for this study. On either side are the inputs: Distance Learning Persistence Studies (left), merged with Graduation and Ranking statistics (right). There were two merge junctures for the data sets from either source. In parallel, graduation statistics were reconciled to be merged into a staging area, which was an Excel spreadsheet. Evidence-based studies were ranked and coded to conform to the statistics, also in an Excel spreadsheet. The two Excel spreadsheets were loaded as flat files from their respective staging areas into BESS1.0, MS SQL Server Express database to be merged (Data Merge) for analysis.

Scenario One: The final merger would consolidate results into a database management system to be compared using queries against the metrics for college completion goals.

Scenario Two consisted of loading data and merging data via query selection.
Scenario One was chosen for this study, as Scenario Two was more work-intensive in a database; however, it could also be possible to merge hybrid scenarios in many database applications. Optimal performance and expeditious handling was the goal of the methodology. At merge junctures, the data were reconciled and prepped for extraction, transformation, and loading into BESS1.0 to be accessed and queried by the data analysis tool.

Figure 2. Best-evidence statistics synthesis (BESS 1.0).

Selected degree attainment data were entered into Microsoft Excel (MS Excel) for staging in preparation for loading into the database. The data were merged from three
sources: The White House Goal 2020 table, the National Council for Education Statistics Bachelor: Degree Attainment for 2014, and the IPEDs survey results for students who are US citizens earning degrees in foreign countries.

A professional database architect validated the selection of the database and the staging of the data. The statistical tools were not warranted in this hypothesis, although they are essential to further research. The reasoning was that effect sizes are primarily used in meta-analysis. Many studies do not report the effect size; however, it can be calculated from other data in the studies for meta-analyses. IBM SPSS is useful; there are also research instruments which will calculate the effect sizes giving the mean and standard deviation. For this study, it was not found to be prudent since the best-evidence studies will be analyzed in future research. A select group of effect size data was taken from 16 blind studies and dummy data were entered for the citation and the abstract.

This suited the purpose of this study to query the data and present it for analysis in a report using the business intelligence tool. The citation and abstract in the studies were used only for sorting and de-duping in this study. The design, hypothetical at the outset, was an examination of the probability to lay the foundation for a novel approach to merge best evidence with disparate graduation statistics for analysis.

**Data Analysis**

The data analysis was conducted in three phases: The Phases were developed and defined Phase I, Phase II, and Phase III. The selected database and the selected data analysis tool were indeed scalable. Based on the longevity of Microsoft it is expected to be able to apply updates or new versions up to and beyond 2025. The intent of this study was to define and construct a data management process which is intact and can evolve for
future analysis. The University of Virginia Library (n.d.) data management resource for researchers was a model for database consistency and conformance. Training and tutorials are readily available from the selected data sources.

**Limitations**

The data gathered consisted of criteria relevant to this study. The criteria were the persistence of new students and adults (25-64) who entered college between 2005 and 2015 in distance education, and who entered college with the intent of attaining a degree. The 284 studies were recorded for further research. Hart (2012) determined much of the difficulty in the literature is due to the lack of consistent terminology in addressing retention, persistence, attrition, and success, or unknown terms used as synonyms or antonyms for persistence. In this study the respective terms used in each study were captured to garner some consistency for future reference. The best-evidence studies consisted of research from 284 studies, primarily within the previous 5 years with some exceptions for reasons noted earlier in this paper. The studies retained are pending exhaustive analysis and coding, they were multi-purposed when populating BESS1.0. The sample size anticipated to increase or decrease during the research process and was limited to a maximum of 300 global studies of findings in distance learning persistence. Some international studies were stored in BESS1.0 for relevance to the long-term research. One international study was included in this study.

The study consisted of two compilations. The first compilation was a meta-analysis type collection of scientific best-evidence distance learning studies on student persistence, both declining and ascending. Studies were stored in the manner depicted in Figure 3. The study is stored by name with the ID number appended as a prefix. The ID
number will always be identical to the record loaded into BESS1.0 which contain the
citation, the abstract and other significant data. In this research the ID number in the
database will match the stored PDF document. The PDF is not stored in the database but
in a separate folder on the hard drive for efficiency.

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSC PLA, Mature Learner.pdf</td>
</tr>
<tr>
<td>RSC National Adult Learner_report_0713.pdf</td>
</tr>
<tr>
<td>RS046 Persistence and adult learner 12 FACTORS166.full.pdf</td>
</tr>
<tr>
<td>RS044 Mullenberg_StudentBarriersToOnlineLearning.pdf</td>
</tr>
<tr>
<td>RS042 Openness and Dropout, a study of 4 university Microsoft Word - Final Paper 262.pdf</td>
</tr>
<tr>
<td>RS040 Qual_Factors Related to DO in OL.pdf</td>
</tr>
<tr>
<td>RS036 Best evidence synthesis alternative to meta anal SLAVIN 1986.pdf</td>
</tr>
<tr>
<td>RS032 Student retention in distance education are we faking our students.pdf</td>
</tr>
<tr>
<td>RS027 Meta Analysis a a choice to improve research.pdf</td>
</tr>
<tr>
<td>RS026 Diplomas and Dropouts final.pdf</td>
</tr>
<tr>
<td>RS025 Efficacy of Hybrid Coursework on Retention in DE.pdf</td>
</tr>
<tr>
<td>RS024 Boston-Adult-Degree-Completion.pdf</td>
</tr>
</tbody>
</table>

**Figure 3.** Best Evidence Research PDFs.

The second compilation was a synthesis of the education databases which were
merged, then loaded to BESS1.0 for analysis in the Business Intelligence application.
Following the two compilations there were attempts to discover a mutually compatible
transformation that would enable a synthesis of the best-evidence statistics with the
education database statistics. Statistical tools IBM SPSS and R were reserved for future
analysis. Combining first the education data sources into one spreadsheet, then the two
compilations, effected a merger, which was found to produce the desired results of
merging the data for analysis. It was determined that the process is better suited to be
carried out in chunks due to the gaps in data in the education database statistics; however, this research laid a foundation for anticipative further analysis of the merged data sets. Per Fink (2005) there are a summary of stages to complete a valid meta-analysis which follows a standard protocol. The missing component in this study was a minimum of two to three researchers to validate the studies using a screening protocol (Fink, 2005). It follows that the best-evidence studies collected should be revisited subject to a minimum of two researches following the standard protocol in further research.

**Summary**

This preliminary research commenced with a collection of graduation statistics from education databases. The methodology continued by including scientific best-evidence based studies. Fink (2005) best practices for meta-analysis compilation was a fortunate discovery during this research. While the full meta-analysis did not occur at this stage it was determined to follow the highest standards and to incorporate them into the knowledge base for future research is of importance. Researchers interested in best-evidence and meta-analyses would be able to easily access the standards and best practices for this approach because it will now be stored in the knowledge base.

The attempt to synthesize the two data elements was made to determine if meaningful data could be elicited. The data from 3 education databases and a sample of best-evidence research articles were selected from the collection of data. The underlying database and the components are expected to evolve into a more sophisticated knowledge base of sets of research specific to certain topics. The knowledge base is expected to benefit preliminary research for future investigators. There are a number of advantages to be gained from a knowledge base. Among them are the ability to benefit from work
performed by others which may not be specialized to their research but is a redundant repetition of steps involved in gathering research. Contributing to the knowledge base would be non-invasive to the content of one’s research, particularly research in progress, and again contributory to the methods of research in general.
Chapter 4: Results

The architecture for this data analysis was depicted graphically, shown earlier in Figure 2. BESS1.0. The database Microsoft (MS) SQL Server Express was installed as the first step. After a thorough review of the databases under consideration it was determined that MS SQL Server Express was appropriate for this research, to expedite the build, securely store, and retrieve the data for analysis. MS SQL Server Express is a free Microsoft version with limited scaling, however, it can be upgraded to a scalable version capable of handling large amounts of data. The free version provided a comprehensive tool for this research. Should the project evolve, there are sufficient offerings within Microsoft and other vendors to scale out or migrate to another database vendor.

Once the database was installed, CA ERwin Workgroup Data Modeler was installed. CA ERwin is known to be compatible with a generous range of vendor databases. CA ERwin is a preferred data modeling tool in the industry due to its capability to port data models from one database to several others databases. CA ERwin also has a reverse-engineering feature which allows for importing a data model from a diversity of databases and re-engineering the design. Although CA ERwin was recommended in the list of applications, it was reserved for the more complex phases of this research. The data collected were reduced to a small-scale prototype making MS SQL Server Express adequate and efficient. The prototype knowledge base was designed in MS SQL Server Express. Eclipse BIRT Report Designer, an open source GUI interface reporting application, hereafter referred to as BIRT, was reviewed for consideration. BIRT was not selected because the goal at this juncture was simplicity and BIRT has a
lengthier learning curve, which is the case with some open source applications. Qlikview, another Business Intelligence application available as a free tool was selected. Qlikview was found to be robust and user-friendly for this research.

Initial steps were carried out to answer the research questions to design and develop a process for the collection, categorization and ranking of evidence-based studies to merge with education database statistics:

**Phase I: Evidence-Based Research**

Question 1.

How can evidence-based research outcomes be ranked and coded for data analysis to merge and measure, in parallel, with traditional statistics and Goal 2020/2025 metrics to determine if there is an impact from declining student persistence outcomes in distance education on college completion rates through 2025?

The collection consisted of 284 evidence-based research papers of various designs in PDF format. From a document storage perspective, PDF was s the preferred format; however, document type is the choice of the researcher. PDF was chosen for its lighter footprint and security features. Less secure document types are not recommended if there are potential inherent security flaws. Web-based studies were converted to PDF format from the Internet to have consistency in the document types on the hard drive. PDF documents were not stored in the database since growth in the pool of documents could affect database performance. In Phase I, the process consisted of searching academic databases for evidence-based studies. The search databases utilized were ProQuest, ERIC, ERIC (EBSCO), Sage, Taylor & Francis, the Lumina Foundation, and Google
There were 284 electronic documents identified through searches for consideration in this study. Several evidence-based studies meeting the search criteria were isolated as a small sample for the merger. All the studies were downloaded in PDF or converted to PDF. The identifying information for each study was entered into the database with an ID number, then subjected to review for inclusion in this research or archiving for future research.

A sample set of blind studies were selected and assigned a pseudo ID number for the merger test. In this test, the actual citation and the abstract were not required to prove the hypothesis, thus, both were replaced with dummy information. Actual statistics data were used from the blind studies to produce real statistics including effect-size, N, and type of study conducted. Codes were assigned based on the type of blind study selected.
These elements were staged in an Excel spreadsheet. The next step was the transformation in preparation to load into the database. The data were sorted by the assigned ID number in the format CC### (C is character, # is a number). The education database records were assigned an ID number using the same format. The ID numbers served as the primary key to join tables in BESS1.0 so analysis could occur.

The electronic PDF documents were stored separately on hard disk, then assigned an ID number identical to the primary key, which would also enhance future retrieval of any PDF document. The search terms used specifically to identify the best-evidence studies were narrowed down to “student retention or attrition, and online or distance learning.” The time frame for collecting evidence-based research was generally held to studies no earlier than five years. There were some older studies (more than 5 years) retained that were included for their archetypal value, or have served as fundamentals used as a basis for investigative analysis. Some of these older studies were retained for justification, provided they were used elsewhere and often in the body of evidence.

These were typically studies which were already mentioned in many of the studies in this area of research (e.g. Tinto, 1975). The studies collected were categorized into sets by a qualifier, such as keywords or a term in a column created for that purpose to enable retrieval of a set of studies relevant to specific searches or a specific topic. It was anticipated this would expedite future research by eliminating some of the redundancy in searching. As new research documents emerge, then the studies which have been touched will have a record in BESS1.0 with pertinent information to assess the value to each study. Additional columns, content or categories could be incorporated, e.g. by research
team, as the need arises. This would not eliminate searches on the typical research databases but enhance it in a knowledge base.

For example, researchers would be able to search by journal, by keyword, by author, by set, in an efficient manner because they would not have to download and touch each study. Researches would be able to benefit from the preliminary findings of other researchers and save valuable time. This would hypothetically mean a reduction in the redundant effort required to search for the same studies, re-examine the same abstracts or statistics only to find they do not meet the purpose of their study. BESS 1.0 was conceived after many years of redundant searching and duplicated effort. The ability to store those efforts into research packages for quick retrieval with prominent research criteria readily available in a knowledge base stored by the standards for valid research, could make the mundane initial tasks reusable and expeditious.

**BESS 1.0 Phase I Process.**

*Raw Collection and Count*

PDF documents, 284 in total, were located utilizing the search criteria student persistence, student attrition or student retention, distance learning or online education.

*Store all PDF documents in a single folder regardless of status*

PDFs were stored in a sources folder, ALL PDFs, regardless of status. Content in other formats was first converted to PDF.

*Identify citations and enter into an Excel spreadsheet*

Citations were identified for every document and were used as a sorting key in the first pass. This was useful to eliminate duplicates in a later step.

*Assign an ID number to each PDF*
An alphanumeric identification number was assigned to each document. This ID was added as a prefix to the file name in the ALL PDF folder. One purpose of the prefix is to control order, and locate the document quickly on the hard drive. Documents were stored remotely on a hard drive rather than in the database.

The PDF document would not be stored in the database: rather in a folder on the hard drive. This method expedites searches and conserves space in the physical database. The researcher can benefit from searches done in prior studies. Researchers can provide their search sets in the knowledge base for future access by other investigators.

*Examples for ID numbers*

- [AR000] Store, not used in this research
- [RS000] US, Used in this research collecting US studies
- [IS000] International used in this research collecting International studies
- [IN000] International, not used in this research

Other IDs can be added later and new categories created.

*Business Rule - Lead with ID, then citation, then abstract when entering data*

This rule was utilized to preserve data integrity during data entry by ensuring there are no duplicates in the collection before entering a new record. All documents are unique by citation; however, the unique ID field serves to enable faster searches. The database engine does not have to read the citation which is a large text field. These two fields should not be concatenated.

*Sort by citation field (author, year, title)*

The citation field was used as a sorting key. Once sorted, those shown to be duplicates were eliminated in the next step.
Dedupe

Delete all duplicate documents. Eliminating duplicates can be automated in a more sophisticated version of this prototype. For this study, manual deduplication was sufficient.

Define Columns for Researchers to Query

Columns (fields) were chosen based on the most accepted standards for valid research methods (Locke et al, 2010).

Assign Primary Category

Primary categories were assigned in a second sweep to delegate each document to a set. Examples of a category type used for clustering sets is below. This is a pseudo example. In practice, more concise codes could be discerned. The category types used were comparative, correlation, effectiveness, effect of, attitudinal

Assign Rank by Category or Keywords

Once the type of study was determined then research documents could be coded by category type in sets or by keywords, if available. The purpose of a rank was to associate the study to a member of a research set. Ways of achieving this could be by extracting keywords from the study, whenever possible. If there were no keywords then rank could be assigned by study type. In this study research was assigned by type of study. Only the effect-sizes, N, and Type are true data. In Table 5, below, two studies were included where N was unknown, for demonstration purposes on the type of data that can be encountered.
ID is the primary key which would be used to join the two tables in the database.

The primary key ID will allow them to be presented in the Business Intelligence Tool interface for analysis. The results from the test are in Table 5.

Table 5

Test Best-Evidence Data

<table>
<thead>
<tr>
<th>ID</th>
<th>Citation</th>
<th>Abstract</th>
<th>EffectSze</th>
<th>N</th>
<th>Code</th>
<th>StudyYr</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS001</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.132</td>
<td>60</td>
<td>Comparative</td>
<td>2014</td>
</tr>
<tr>
<td>RS002</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.171</td>
<td>155</td>
<td>Comparative</td>
<td>2012</td>
</tr>
<tr>
<td>RS003</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.264</td>
<td>110</td>
<td>Comparative</td>
<td>2014</td>
</tr>
<tr>
<td>RS004</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.373</td>
<td></td>
<td>Effectiveness</td>
<td>2013</td>
</tr>
<tr>
<td>RS005</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.570</td>
<td>88</td>
<td>Comparative</td>
<td>2013</td>
</tr>
<tr>
<td>RS006</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>1.113</td>
<td>50</td>
<td>Effect Of</td>
<td>2013</td>
</tr>
<tr>
<td>RS007</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.281</td>
<td>53</td>
<td>Effect Of</td>
<td>2013</td>
</tr>
<tr>
<td>RS008</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.028</td>
<td>168</td>
<td>Attitudinal</td>
<td>2012</td>
</tr>
<tr>
<td>RS009</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.800</td>
<td></td>
<td>Comparative</td>
<td>2014</td>
</tr>
<tr>
<td>RS010</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.920</td>
<td>9</td>
<td>Comparative</td>
<td>2014</td>
</tr>
<tr>
<td>RS011</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>-0.420</td>
<td>512</td>
<td>Correlation</td>
<td>2014</td>
</tr>
<tr>
<td>RS012</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.078</td>
<td>344</td>
<td>Correlation</td>
<td>2014</td>
</tr>
<tr>
<td>RS013</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.464</td>
<td>147</td>
<td>Observational</td>
<td>2012</td>
</tr>
<tr>
<td>RS014</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.078</td>
<td>344</td>
<td>Correlation</td>
<td>2013</td>
</tr>
<tr>
<td>RS015</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.464</td>
<td>147</td>
<td>Comparative</td>
<td>2014</td>
</tr>
<tr>
<td>RS016</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>0.104</td>
<td>73</td>
<td>Comparative</td>
<td>2014</td>
</tr>
</tbody>
</table>

Phase II: Education Databases

The next step, designated Phase II, the collection, then extraction, transformation and loading of the education data was performed to answer the research question shown below.

Question 2

How can statistics for college completion from 2005 to 2015 be disaggregated to compare, in parallel, with evidence-based studies in an analysis to consider declining student persistence outcomes in distance education from 2005 to 2025?
“The White House: Meeting the Nation’s 2020 Goal: State Targets…” was selected as the baseline for the merger, initially, looking ahead to a future longitudinal study between 2005 to 2025. Since the purpose of this study is how to merge the data, the research was not expected to result in significant statistics, nor to answer whether student retention was declining or ascending. Rather, the success of this methodology was an infrastructure to perform the merger and for that infrastructure to be suited for replication and for long term analysis. The education datasets, hereafter referred to as Phase II, were researched to identify samples for this study. The National Center for Educational Statistics standard was examined to understand the structure of the various datasets. The research also consisted of viewing tutorials on the education databases named below. It is recommended that researchers, particularly for a study including disparate education databases view the tutorials on the education sites. This is where a more succinct comprehension of the gaps in the education databases is revealed.

The education databases selected were narrowed down to those in the National Center for Education Statistics (NCES): Distance Learning Datasets. Several datasets were examined: National Postsecondary Student Aid Study (NPSAS), Beginning Postsecondary Students Longitudinal Study (BPS), Baccalaureate and Beyond Longitudinal Study (B&B), and Postsecondary Education Sample Survey Datasets. The objective was to understand how to measure results from the NCES state level to compare, over time, the resulting graduation statistics and any hidden data, as relates to Goal 2020/2025. Because the education datasets conform to a standard, one sample dataset would suffice to answer the research question. Additionally, the data structures are well documented to examine any issues going forward.
Table 104.88 was selected as the sample dataset which was created by extracting only the bachelor degree and above attainment for persons’ age 25 and over, excluding those columns broken out by sex and state for 2014 data. Only the highlighted data Column 2 “Total degrees (in thousands) for persons’ age 25 and older” by state including the total figure for the United States appearing at the top in this spreadsheet and Column 8, the “Total column for “Percent with bachelor’s or higher degree” were extracted for this merger. The data in part is shown in Figure 4.

Bachelor’s degree or higher data were extracted and staged in a new Excel spreadsheet. Additional columns were added. Excel formulae were used to transform the percentages to numeric data, and to transform the Total figures by state to figures in thousands. For this study the typical stages in creation of a database from disparate sources consist of extraction, transforming, and loading (ETL) into the new database, the knowledge base.
ETL was performed manually for demonstration purposes. Typically, ETL can be easily automated and is usually done so for large datasets. A new Excel spreadsheet named (Education Database) EDDB.xls was created using the extracted data. Columns were added to affect the merger. The new columns were:

- **ID** – Unique ID number and primary key
- **Database** – Database name
- **TblName** - Education dataset name
- **DateAdded** - Date added to BESS1.0
- **Year** – Year of the study data
- **StCt** – State numeric count of degrees (transformed to thousands)
- **StPctTtl** – Total percent for the State
- **StPctUs** – State percent of the US total
- **BDegPctSt** – Bachelor degree percent of State degrees
- **BDegCtSt** – Bachelor degree numeric count of State degrees (transformed to thousands)
- **PctValidation** - Formula to validate the accuracy of the count

Under the auspices of the Institute for International Education (IIE), Belyavina & Bhandari (2012) conducted an international survey by country. They collected counts for U.S. students in overseas degree programs in. Only the figures for students earning bachelor degrees were extracted. This data were never collected in the past (Belyavina & Bhandari, 2012), making the survey data an excellent candidate for hidden graduation statistics data. Table 6 contains the data for US students who earned bachelor’s degrees abroad taken from the survey (Belyavina & Bhandari, 2012). There were additional
counts for students earning master’s and doctoral degrees in foreign countries. In this study, only the bachelor’s degree data were extracted. Country level data were merged onto the spreadsheet with state level data, shown in blue. The degree counts were made numerically consistent once the state level dataset was transformed to thousands so they could be added to the country level counts and provide an account of the degrees attained for attainment between 2009 and 2014.

Table 6

**U.S. Undergraduate Students Earning Degrees Abroad**

<table>
<thead>
<tr>
<th>Country</th>
<th>Undergraduate</th>
<th>UG Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1099</td>
<td>38%</td>
</tr>
<tr>
<td>China</td>
<td>875</td>
<td>53%</td>
</tr>
<tr>
<td>Denmark</td>
<td>39</td>
<td>27%</td>
</tr>
<tr>
<td>France</td>
<td>1581</td>
<td>47%</td>
</tr>
<tr>
<td>Germany</td>
<td>649</td>
<td>17%</td>
</tr>
<tr>
<td>Ireland</td>
<td>460</td>
<td>48%</td>
</tr>
<tr>
<td>Japan</td>
<td>283</td>
<td>51%</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1499</td>
<td>82%</td>
</tr>
<tr>
<td>UK</td>
<td>5940</td>
<td>37%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12425</strong></td>
<td></td>
</tr>
</tbody>
</table>

Undergraduate figures extracted from Institute for International Education study Belyavina & Bhandari (2012)

Three disparate data sources were merged in the one table. The White House goals were set using data collected in 2009. The IDs for country level data was prefixed with CC, while the state level IDs were prefixed with ST.

A range of years was chosen in this instance (2009-2014) merging chunks of data in closest proximity to do analysis on this chunk of data. Other chunks could be defined in a new set (e.g. 2005 – 2008; 2015 - 2017). Qualifiers (CC###, ST##) were for identification and integration so the data could be merged. Country level IDS (CC001 – CC010) could be stored in the same column as State level data (ST001- ST0050), as the
counts were for bachelor’s degrees attained in the same year. While the state level data was identified as bachelor’s degrees and above, this example serves for demonstration purposes. In Figure 5 is a screenshot of a sample from the White House 2020 goals.

![Figure 5](image)

**Figure 5.** Screenshot of sample state goals for White House 2020


The data above is archived per the Department of Education; however, there were updates released by The Lumina Foundation on 2017 figures which calculated 16 million additional degrees or credentials required to attain Goal 2020/2025.

The counts and percentages were extracted from this spreadsheet and merged onto the Excel staging spreadsheet. In this research the “Graduates as of 2009” were used as the baseline for the projections “Graduates as of 2020.” The figures for bachelor degrees were merged onto a single Microsoft Excel spreadsheet, along with the Institute for International Education survey figures for US citizens earning degrees abroad (Belyavina & Bhandari, 2012), utilizing totals by country.
The resulting totals were then merged with the baseline of the 2009 degrees completed from the White House table for the Nation’s 2020 goals by State. Multi-level headers were removed since these can cause anomalies when loaded into the database. New headers were created on the spreadsheet suited for this research. All header names were defined in a Data Dictionary.

Aggregate gaps exist in the education databases, as predicted, due to aggregation and decisions made in the past for various reasons resulted in the omission of data collection for some years. The gaps are well documented in the National Center for Education Statistics modules and will be addressed in more detail in further research. Future research will also elicit a decision for the most efficient automation of this approach. There were also findings in the research noted earlier in this paper, relevant to Cook and Pullaro (2010) of American Council of Education (ACE), who noted there are inconsistencies in the 1995 6-year graduation rates between the National Center for Education Statistics (NCES) and the Integrated Postsecondary Education Data System (IPEDS). It will require further study to ascertain if this inconsistency continues because of disparate data collection procedures. Some data are collected in a limited capacity in the Integrated Postsecondary Education Data System (IPEDS).

Due to inconsistencies in the data collection, the process above sets the stage for further study to collect several sets of available datasets with distance learning components and attempt the hypothesis for a select year by looking for a chunk of data that would examine datasets and studies in a range across a selected year or years. To determine if the standalone data structure could be staged and merged, the decision was made to opt for a specified year, plus or minus 2 years, from the available Distance
Learning Dataset closest to the White House goal and the international data which was 2012, then attempt to merge data in that range (2009–2014) across datasets. The idea of chunking to analyze disparate data was borrowed from research in psychology. There is an abundance of cognitive research on the benefits of chunking primarily focused on retention or recall in cognition. Gilchrist (2015) conducted research on Chunk Measurement Methods which, inadvertently, appeared to offer potential for analyzing disparate data and eliciting commonalities to be reconstructed later utilizing coding and relationships, a qualitative approach.

Each of the methods to measure chunks that are discussed share a notable commonality—each is based upon a fundamental property of chunks. For instance, it is presumed that items that are a part of the same chunk are tightly-bound or compressed (Gilchrist, 2015).

Hays and Singh (2012) used the qualitative methodology: chunking, coding and reducing mnemonic data into smaller sets. Coding was a small part of their qualitative study. The application of chunking, chunking measures, and coding could potentially offer new insights if found to be useful in reducing discrepancies between education databases. To illustrate the procedures Hayes and Singh (2012, p. 295) referred to a grounded theory qualitative study of counselors’ decision-making processes by Hays, McLeod, and Prosek (2009). Notwithstanding the outcome of the study, the results were a sequence from “a priori data reduction, to several iterations of analysis, to data management as a tool in collection and analysis” (Hayes & Singh, 2012, p. 295), making the chunking and coding methodology of great interest to future research for this study.
Beyond that “The Institute of Education Sciences (IES) uses Restricted-use data licenses as a mechanism for making more detailed data available to qualified researchers.” Restricted-use data may also provide more detail, thereby closing even more gaps in education datasets, given a non-invasive method to examine the datasets. This would be a consideration for future research.

**Phase III - Merging the Data**

Phase III consisted of loading the data into the Microsoft SQL Server Express to test the ability to query the data from the two sources. The uploaded data were a combination of actual data and test data with a dummy component for the abstract and citation columns from the best-evidence studies collected. The abstracts and citations have been established in this study as useful for sorting and deduping the PDFs collected and stored. The abstract and citation content, which are text fields subject to limited characters when in use, were not needed to test the hypothesis and were excluded for better performance on the database. The hypothesis test was not expected to resolve into results of significance in the data. Using actual statistical data from studies for valid figures with pseudo labels, the objective was to test if it was possible to merge the data to and produce viable results. To achieve significance the process should be repeatable indefinitely, unless or until there are changes to any data structure requiring modification in the database. In the event of changes to the data structure this can and should be resolved and updated in BESS1.0 by a database professional. BESS1.0 is a relational database used to query and analyze data using SQL or an analysis tool. If the data stored should exceed terabytes, then alternative database options should be considered. This is not yet an approach for Big Data databases. In the rapid escalation of information and
technology there should be a future for BESS1.0. Although worthy of mention, Big Data is outside the scope of this study. Understanding how to merge the data sources and generate a report, however, opens possibilities for future research.

Processing could occur once the synthesis was complete. The measures considered from the education datasets were counts, percentages, totals, outcomes the results from education database statistics best suited for synthesis.

Chunks of data were tested then examined for accuracy. Once synthesis was achieved the data could be analyzed and queried for viewing and analysis using the business intelligence and data analysis tool. The process described above was enacted to answer the following questions:

**Question 1**

How can statistics for college completion from 2005 to 2015 be disaggregated to compare, in parallel, with evidence-based studies in an analysis to consider declining student persistence outcomes in distance education from 2005 to 2025?

**Question 2**

How can statistics for college completion from 2005 to 2015 be disaggregated to compare, in parallel, with evidence-based studies in an analysis to consider increases in student persistence outcomes in distance education from 2005 to 2025?

To answer the above research questions, Department of Education state goals compiled for the White House for Goal 2020/2025 were used as the basis for achievement benchmarks on the state level. Data from sources sponsored by the Lumina Foundation are expected to be reliable for figures, going forward, on attainment.
outcomes. It is uncertain whether the White House will follow through with Goal 2020; however, the Lumina Foundation’s Goal 2025 is robust and continues. On the education database side the National Center for Education Statistics Attainment data sets are expected to continue as a viable source. Also, enrollment percentages for the national and regional state levels captured annually by National Student Clearing House Coverage: Enrollment Coverage is an annual figure for measuring declining student persistence.

Question 3

How can the metrics for Goal 2020/2025 be quantified to integrate declining student persistence outcomes in distance education to ensure there are no “aggregation gaps” (Horowitz, 2014) in the percentages for college completion due to declining student persistence between 2005 and 2025?

The primary purpose of the ranking and coding process in this research was to achieve the mapping and synthesis for this study. The three education datasets were merged into one. A small sample of blind best-evidence research was captured in another spreadsheet using actual effect sizes. The work to identify which were valid studies and to capture the effect sizes with identifying features was not the goal of this research. The goal was to test how these disparate sources could be merged. The merger was a success in the infrastructure of BESS1.0. That it can be collected and made to integrate was the first step and that first step has succeeded.

A secondary benefit was anticipated. Each study was ranked and coded to a category or subject matter, even if not included in this research. The secondary benefit was expected to be a search methodology that would reduce the redundancy of searching. The same set of studies for similar research topics or repeat searches already completed
by individuals or groups of researchers would be stored for retrieval in sets. By capturing
the work performed, if this results in, at minimum, a re-usable database of searches for
studies, categorized in sets, it could be useful for saving time and effort.

Question 4

How can evidence-based research outcomes be ranked and coded for data
analysis to merge and measure, in parallel with traditional statistics and Goal
2020/2025 metrics to determine if there is an impact from declining student
persistence outcomes in distance education on college completion rates through
2025?

The goal to rank and code evidence-based research studies was accomplished
without the necessity to include or compile the details of those studies to meet the
hypothesis. The coding for this study was developed for identifying and tracking the
studies within BESS1.0 and was not intended to be identical to the coding methodology
as defined by Singh (2012), although that qualitative coding may be useful as a guide,
later as BESS1.0 is expected to evolve. The infrastructure in this study is more simplistic
for test data. The data expected to be collected for analysis in the future is expected to be
larger in content; however, it should follow the same structure as the data used in this
study. Thus, the criteria, that the database selected must be scalable or updatable over the
long term, enable broad capabilities for data elements, regardless of the size of this
sample. Alternately there could be a tool (CA ERwin) capable of porting to several
relational databases. (See Figure 6)
Figure 6. Disparate Data loaded into CA Erwin® Data Modeler. Screenshot CA Erwin®.

The ranking and coding developed for this study has succeeded in categorizing the elements of significance in best-evidence research: effect-size, N, Type of Study, for analysis. In combination with the chunks of data from education databases, the effect-sizes of best-evidence research studies can be analyzed to determine which research elements in distance learning are most conducive to student attrition and degree attainment sought for Goal 2020/2025. A simple dashboard was created in Excel that does not require any additional applications nor changes to existing technology. A tutorial would and should also accompany BESS1.0 as an outcome of this research. The
BESS1.0 Best-Evidence Research and Education Database Degree Attainment dashboard (Figure 7) is below.

![Dashboard Image]

**Summary**

The data from disparate sources identified earlier in this chapter were extracted from the original datasets or from best-evidence studies, then staged in two Excel spreadsheets. The data were transformed in each of the staging spreadsheets to be uploaded to a database for analysis. Excel formulae were used for validation and transformations. The two Excel spreadsheets were converted to .csv files for ease of uploading into MS SQL Server Express.
The final step was to create a dashboard as the outcome for this study. The BESS1.0 Dashboard was more of a challenge than anticipated. It was, however, a good exercise to grasp the complexity of the learning curve for non-technical researchers. The build was a data architecture endeavor. The tool must be friendly for those who are non-technical. Several attempts were made with various tools to create a dashboard. There were repeated obstacles having nothing to do with research. There are many options for creating reports and dashboards. For some a report, will suffice and a dashboard typically falls into the realm of a nice to have for the average user, despite an abundance of technology. There would be sufficient choices suited to researchers at various technical skill levels. It is worthy of note that even with very advanced skills there were obstacles which had nothing to do with capability in this study. Among them were incompatibility of operating system, versions of supporting applications, such as Microsoft .Net or Windows.

There were 284 best-evidence research studies collected to demonstrate the collection of research for this study. The studies were converted to PDF as needed, and assigned an ID appended to the title. The studies were recorded in a spreadsheet and sorted, de-duplicated, and either archived for later studies or selected for a more in-depth study of best-evidence student retention. Several education databases and surveys were reviewed for statistical data. Three sources were used in this research by recording in a spreadsheet. Both sources were transformed in preparation for loading into the database. Once loaded, they were refined to run in queries for analysis. Several reporting and analysis tools were tested. There were too many inconsistencies with the computer, the version of the operating system, and the versions of applications that could be configured
on the existing system. A clean dedicated server is recommended with system requirements to meet the desired applications. The final step to create a report, then a dashboard was completed in Excel for demonstration purposes. Notwithstanding, the obstacles with technology, the research questions were answered successfully. The relevant steps were demonstrated in screenshots and each goal was accomplished successfully. There is also far-reaching prospect to diversify and expand on this simple approach.
Chapter 5: Discussion

Summary of Findings

The most significant finding in this study was the survey conducted by Belyavina & Bhandari (2012) that uncovered the figures for U.S. students earning degrees abroad. That these figures added even modest counts to the 16 million additional credentials appears to justify the continuation of this research. Haydarov, Moxley & Anderson’s (2013) late findings in the document collection examined a hypothesis that suggest lowers retention rates are indicative the institution is performing poorly. Their study was of great interest and archived for further review. It was of great interest uncovering that Haydarov et al. (2013) and many others met with obstacles commencing the research because they had to reconcile the many interchangeable semantic variations to concede on what was attrition, retention, and other terms relevant to their research. It is worthy of mention that Haydarov, et al. also found that retention and attrition are synonymous; however, used differently in education for student retention and institution attrition, respectively.

The Lumina Foundation (2017) updated its Strategic Plan reporting that an additional 24.2 million Americans will earn postsecondary credentials. While the numbers, at first glance, seem generous and promising, it will still require an additional 16.4 million degrees or credentials to reach Goal 2020/2025 (2017). As suggested in this study, there may be more numbers and more sources which will buffer the numbers, despite figures on declining online graduation statistics. The macrocosm may be a global degree attainment melting pot which holds a modest wealth of uncaptured data.
This study identified the Goals of the White House in the previous administration and the Lumina Foundation to increase the numbers of US citizens who have attained degrees, certifications, or credentials by 60% by 2020 and 2025, respectively. Despite changes in the political climate, thanks to the efforts of the Lumina Foundation and the partners who committed to the plan, Goal 2020/2025 is from all perspectives still intact. The sources of student education data including graduation statistics and degree attainment were researched and answered the research questions. In doing so, the door has been opened to further refine the knowledge base and proceed to the relevant research on best-evidence.

This study focused on the distance learning component, seeking how to uncover “aggregate gaps” in the disparate data collected to understand if there were significant numbers of students who simply fell through the cracks. There may be opportunities ranging from modest to vast, untapped sources and in the graduation statistics for Goal 2020/2025. There were discrepancies in data collection methodologies, and content between education datasets, such as the National Council on Education Statistics and the Institute for Postsecondary Education Datasets, as well as gaps in which year data were collected making a standard longitudinal unlikely. Data is, to date, not captured in consecutive years. The reasons are not always apparent. It requires intense research to grasp the reasons for the disparities in the education databases. Some of the reasons data back to years when data collection was purely paper-based. The relationship between early technology, or the lack thereof, and the standards in place today needs to investigated further.
The study consisted of data collected from best-evidence research, and datasets from education databases. The data were reviewed in a meta-analysis approach, exclusively, for integration, rather than performing a meta compilation of best-evidence studies or education databases. It is an expectation that once this approach, how to achieve the merger, was determined to be a success, then a detailed meta approach could proceed in a later stage of this research per the standards and best practices for meta-analytic best-evidence compilation. There is also consideration for a broad expansion of the data that can be collected and stored in BESS1.0. Given a more sophisticated architecture that would be thoroughly analyzed to scale to meet research up to and through 2025 the idea of BESS1.0 presents potential for a rich application. Funding of course would be required to proceed to the sophistication that can be achieved. There are experts capable of designing an architecture suited for Big Data, the latest trend on the bleeding edge in higher education.

CA ERwin was acquired from erwin, and examined for the capabilities which would enhance this architecture. CA ERwin capabilities would be essential in later stages of this research for reverse engineering the database and for portability to another vendor, if needed. It is also a preferred industry tool for data architects and data modelers.

The samples were three education datasets: (a) The White House 2020 Goals by State set in 2009, (b) The Institute for International Education Survey: US Citizens Earning Degrees Abroad in 2012, (c) the National Center for Education Statistics Bachelor degree and above, attainment for students 25 and over in 2014. The education data represents a range of data in closest proximity to make sense of the merger. While it is not known whether the White House will continue Goal 2020, it is inherent within the
Lumina Foundation’s Goal 2025. For historical purposes the White House goals were the basis for the comparisons. It is expected that the goals will remain intact through changes in administration between now and the culmination of Goal 2020/2025. The Institute for International Education survey data consists of “aggregate gaps” (Horowitz, 2015) in bachelor’s degree attainment by US students earning degrees abroad. The data were merged into one spreadsheet and uploaded into BESS1.0.

Best-evidence studies collected using the search terms described earlier. Real effect-size statistics were elicited from 16 blind studies from the body of studies collected and stored in PDF format to the hard drive (rather than in BESS1.0). The purpose was to maintain better performance in the MS SQL Server database. The PDF documents were BESS coded with an ID number to keep track of the PDF files on the hard drive for future access.

To ensure validity of a meta-analysis collection the studies used should be based on a standard protocol. Fink’s (2005) conclusion that effect sizes based on larger studies have more stability was evidenced even in the pseudo data from best-evidence studies collected that did have real data but pseudo labels. The conclusion (2005) that larger studies should be weighted more heavily than smaller studies is a pattern to anticipate in further research. To elicit the effect-sizes by conducting a systemic literature review per the standard protocol would best serve this study if conducted in further research.

**Interpretation of Findings**

The study resulted in an express need for sharing research and BESS1.0, a new small-scale but broadly scalable knowledge base that enables research teams to collaborate and reduce redundancy. Researchers can store, categorize, track, and recover
the studies they have touched in a central location. New members of a research team can independently and quickly get up to speed on the current state of a research project by utilizing BESS1.0 to learn the status of a project. Universities, research organizations, and graduate students will find BESS1.0 of interest in enhancing the quality of their research. Additional proprietary datasets can be loaded into BESS1.0. The current study in no way limits the utilization of BESS1.0, and BESS1.0 has no limitations for topics, data sources, data structures, nor data capture except for limitations inherent within a specific database. The solution is simply to choose a different database, as the infrastructure for BESS1.0 can be applied across various relational databases.

**Limitations**

To achieve the synthesis Microsoft SQL Server Express was installed on a desktop server. It is recommended that the installation have a dedicated server for this process, in practice, for security and performance. Microsoft SQL Server Express was well suited for research, exploration, and testing. There were other databases considered which would also have served well, or perhaps better. Those were not suited for the rapid development and familiarity with the technology required in this particular study. In a real-world application Microsoft SQL Server Express would not ordinarily be considered as a candidate database beyond prototyping. Other limitations are funding to develop the knowledge base to the desired level of sophistication, including compensation for the best in class expertise.

**Implications of Findings**

Several organizations were approached for funding for this research over the past year or more. It is noteworthy that the search for funding has suggested the importance
of the topic of this study in the research industry. In lieu of having a funding source, or even if having a funding source, a grant and foundation component with history may also be of service to investigators. Statistical information and best practices could be included in the knowledge base, also a tutorial library could be included. The possibilities are endless, all dependent on the content, which would be useful for one-stop shopping to the research community or a limited community of fellows, for example. There could also be levels of access to information in the knowledge base. It will require more in-depth analysis and the skills of a dedicated data architect and software engineer to realize the complete vision of BESS1.0 (Best-Evidence Statistics Synthesis) application.

**Recommendations for Further Research**

It is recommended that researchers continue identifying degrees, certificates and credentials overlooked in degree attainment (Belyavina & Bhandari (2012). As noted earlier, the Lumina Foundation (2017) reported that over 16 million degrees are still needed to reach Goal 2020/2025. It is expected, that the figure may already be modified or soon will be due to the ongoing dynamic nature of Goal 2020/2025. It is not known whether the figures for the IIE were counted in the 2015 update where the Lumina Foundation reported that 16 million degrees are still needed.

**Summary and Conclusion**

Conducting this study was an expose of the many intricate avenues to be investigated on student persistence. As it stands, there is still no common terminology, there is no consistent data collection, there is no preferred, or tried and true methodology. The history of education data collection is fraught with changes, updates, and additions while simultaneously retaining procedures that have been questioned for validity. The
origin of this study was motivated by the dilemma that policy makers, academia, politicians, funding and social services continue to rely on inaccurate student data to make important decisions. For this model to continue would be egregious. This must be corrected, and even if incrementally; updated to better meet the solutions sorely needed to advance degree attainment and student persistence.

The conclusion of this research leads to great anticipation to further develop BESS1.0 beyond this simplistic prototype to a more sophisticated application. The nature of research suggests that BESS1.0 should be maintained, at least initially, in a non-cloud, secured environment on a private network where collaboration can occur without risk to the research. BESS1.0 can be scaled out to securely investigate aggregate and detail data. Reference data (e.g. postal codes, country codes, academic institutions both national and global) can be investigated by links in a less secure web environment. As a knowledge base BESS1.0 would ideally store content which is relevant to the investigators current research in a one stop shopping view. As needs arise, new content could be made available in a myriad of possibilities: links, PDFs, colleagues, other investigators, peers, even a chat room in a less secure environment. It would be prudent to have levels of security or if desired turn off less secure levels. It is of course important to continue the process of discovering “aggregate gaps” in graduation statistics and closing the gap to Goal 2020/2025.
References


National Science Foundation Engineering (ENG) data management plan requirements.
Retrieved from https://www.nsf.gov/eng/general/dmp.jsp

National Student Clearinghouse Research Center (NCS). Retrieved from
https://nscresearchcenter.org/workingwithourdata/


Russo-Gleicher, R., (2014). Improving student retention in online college classes:
Qualitative insights from faculty. Retrieved from
http://csr.sagepub.com/content/16/2/239.abstract

http://www.workingoutwhatworks.com/en-GB/Magazine/2014/10/Statistics_in_educational_research


Appendix A

Tinto’s Model
Tinto’s Model

Appendix B

Mobilizing to Reach Goal 2025
The Lumina Foundation Strategic Plan *Mobilizing to Reach Goal 2025* from The Lumina Foundation. http://www.luminafoundation.org
Appendix C

EDDB Microsoft Excel Compilation
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>3259</td>
<td>3426</td>
<td>3613</td>
<td>3814</td>
<td>4021</td>
<td>4238</td>
<td>4463</td>
<td>4700</td>
<td>4952</td>
</tr>
<tr>
<td>Alaska</td>
<td>440</td>
<td>460</td>
<td>480</td>
<td>500</td>
<td>520</td>
<td>540</td>
<td>560</td>
<td>580</td>
<td>600</td>
</tr>
<tr>
<td>Arizona</td>
<td>1433</td>
<td>1476</td>
<td>1521</td>
<td>1570</td>
<td>1622</td>
<td>1676</td>
<td>1732</td>
<td>1790</td>
<td>1850</td>
</tr>
<tr>
<td>Arkansas</td>
<td>1970</td>
<td>2013</td>
<td>2056</td>
<td>2100</td>
<td>2145</td>
<td>2191</td>
<td>2238</td>
<td>2286</td>
<td>2336</td>
</tr>
<tr>
<td>California</td>
<td>2048</td>
<td>2095</td>
<td>2143</td>
<td>2192</td>
<td>2241</td>
<td>2291</td>
<td>2343</td>
<td>2396</td>
<td>2450</td>
</tr>
<tr>
<td>Colorado</td>
<td>2187</td>
<td>2235</td>
<td>2286</td>
<td>2338</td>
<td>2390</td>
<td>2443</td>
<td>2498</td>
<td>2554</td>
<td>2611</td>
</tr>
<tr>
<td>Connecticut</td>
<td>2410</td>
<td>2463</td>
<td>2518</td>
<td>2575</td>
<td>2633</td>
<td>2693</td>
<td>2755</td>
<td>2820</td>
<td>2886</td>
</tr>
<tr>
<td>Delaware</td>
<td>2426</td>
<td>2483</td>
<td>2543</td>
<td>2605</td>
<td>2669</td>
<td>2735</td>
<td>2803</td>
<td>2873</td>
<td>2945</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>666</td>
<td>690</td>
<td>716</td>
<td>743</td>
<td>772</td>
<td>802</td>
<td>834</td>
<td>868</td>
<td>904</td>
</tr>
<tr>
<td>Florida</td>
<td>1032</td>
<td>1069</td>
<td>1108</td>
<td>1149</td>
<td>1192</td>
<td>1238</td>
<td>1287</td>
<td>1339</td>
<td>1394</td>
</tr>
<tr>
<td>Georgia</td>
<td>4578</td>
<td>4876</td>
<td>5197</td>
<td>5540</td>
<td>5906</td>
<td>6295</td>
<td>6698</td>
<td>7115</td>
<td>7545</td>
</tr>
<tr>
<td>Hawaii</td>
<td>874</td>
<td>916</td>
<td>961</td>
<td>1009</td>
<td>1059</td>
<td>1113</td>
<td>1169</td>
<td>1228</td>
<td>1290</td>
</tr>
<tr>
<td>Idaho</td>
<td>1089</td>
<td>1140</td>
<td>1195</td>
<td>1253</td>
<td>1314</td>
<td>1378</td>
<td>1444</td>
<td>1513</td>
<td>1585</td>
</tr>
<tr>
<td>Illinois</td>
<td>1065</td>
<td>1124</td>
<td>1185</td>
<td>1248</td>
<td>1314</td>
<td>1382</td>
<td>1452</td>
<td>1525</td>
<td>1600</td>
</tr>
<tr>
<td>Indiana</td>
<td>3044</td>
<td>3264</td>
<td>3496</td>
<td>3739</td>
<td>3995</td>
<td>4265</td>
<td>4549</td>
<td>4846</td>
<td>5155</td>
</tr>
<tr>
<td>Iowa</td>
<td>2063</td>
<td>2145</td>
<td>2239</td>
<td>2336</td>
<td>2435</td>
<td>2538</td>
<td>2645</td>
<td>2756</td>
<td>2871</td>
</tr>
<tr>
<td>Kansas</td>
<td>1891</td>
<td>1987</td>
<td>2087</td>
<td>2192</td>
<td>2300</td>
<td>2412</td>
<td>2530</td>
<td>2652</td>
<td>2780</td>
</tr>
<tr>
<td>Kentucky</td>
<td>2969</td>
<td>3113</td>
<td>3261</td>
<td>3414</td>
<td>3572</td>
<td>3736</td>
<td>3905</td>
<td>4080</td>
<td>4260</td>
</tr>
<tr>
<td>Louisiana</td>
<td>2063</td>
<td>2145</td>
<td>2239</td>
<td>2336</td>
<td>2435</td>
<td>2538</td>
<td>2645</td>
<td>2756</td>
<td>2871</td>
</tr>
<tr>
<td>Maine</td>
<td>937</td>
<td>987</td>
<td>1040</td>
<td>1096</td>
<td>1155</td>
<td>1217</td>
<td>1282</td>
<td>1350</td>
<td>1420</td>
</tr>
<tr>
<td>Maryland</td>
<td>4238</td>
<td>4463</td>
<td>4700</td>
<td>4952</td>
<td>5218</td>
<td>5498</td>
<td>5790</td>
<td>6104</td>
<td>6432</td>
</tr>
</tbody>
</table>

EDIES Microsoft Excel 2016 Database Compilation (UMHO). US Students in Overseas Degree Programs (Belayoun & Bhandari, 2013): Meeting the Nation’s 2010 Goal/State Targets for Increasing the Number and Percentage of College Graduates With Degrees: National Center for Education Statistics: All Levels of Education by Educational Attainment. Table 104.88 from nces.ed.gov.