I, Aaron D. Profitt, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Educational Studies.

It is entitled:
Transparent, Accessible Accountability in Higher Education: A Sector-focused Approach

Student’s name: Aaron D. Profitt

This work and its defense approved by:

Committee chair: Mary Brydon-Miller, Ph.D.
Committee member: Vicki Plano Clark, Ph.D.
Committee member: Christopher Swoboda, Ph.D.
Transparent, Accessible Accountability in Higher Education: A Sector-focused Approach

A dissertation submitted to the
Graduate School of the University of Cincinnati

In partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

In the School of Education
of the College of Education, Criminal Justice, and Human Services

2015

by

Aaron D. Profitt
MA, University of Kansas

Committee chair:
Mary Brydon-Miller, Ph.D.
Abstract

U.S. higher education institutions participate in a number of transparency initiatives intended to enable stakeholders to hold the institutions accountable. In these initiatives, and the broader higher-education accountability literature, there are two gaps: (1) accreditation agencies and processes are generally not included and (2) evidence does not exist to demonstrate stakeholders’ ability to interpret, and their satisfaction with, transparency systems. Working within a single U.S. higher-education sector (Biblical higher education) as defined by an accrediting agency (the Association for Biblical Higher Education), this three-phased research project developed a system whereby institutions of higher education may hold themselves accountable to a broad range of stakeholder groups in a transparent fashion, satisfying stakeholders’ information needs. Phase One surveyed accreditation experts to develop a multivariate index of institutional health; this phase included identifying metrics for inclusion in the index and identifying cutpoints to classify an institution as at-risk, challenged, effective or excelling. Results were subjected to validation by experts unaffiliated with the accrediting agency; validation did not conclusively support the developed index, reflecting the tension identified in the literature between improvement- and accountability-oriented paradigms. Phase Two surveyed a broad range of internal and external institutional stakeholders to determine what kinds of information they perceived themselves to need in order to evaluate a higher-education institution’s performance. Based on those results, a website template based on Phase Two’s survey results was created, and four mock profiles were created. Phase Three surveyed the same broad range of stakeholders as in Phase Two; here, stakeholders were asked to review the mock online profiles, then answer questions measuring (a) the accuracy of their understanding of profile information and (b) their satisfaction with the profiles as an accountability mechanism. Accuracy, which differed
significantly across stakeholder groups, was moderate, suggesting that transparency initiatives may not empower stakeholders, particularly consumers, as much as is generally assumed. Satisfaction was mediocre, not differing significantly from a “neutral” response, further challenging an easy, *prima facie* belief in transparency initiatives’ value. In addition to demonstrating methodologies that may be used in sector-focused accountability initiatives, results raise serious questions about stakeholders’ understanding and interpretation of common accountability metrics, providing both further research direction and action opportunities for institutions and agencies.
Acknowledgements

No one other than my wife, Lisa, fully understands all that has gone into this degree and dissertation, and no one else deserves my thanks as much as she. Thank you, Lisa, for your encouragement, input, counsel, perspective and love throughout these years! Thank you for helping our sons to celebrate this work with us, too.

My parents should also be thanked publically here, as they raised me immersed in education, including in the atmosphere of a Bible college. Little did they know that I would someday complete a dissertation dealing with Bible-college education! Thank you, Dad and Mom, for cheerleading for me during my work on this degree.

I am deeply grateful for the financial and workload-reduction support from God’s Bible School and College, especially from its leader, Pres. Michael Avery, and my mentor, Dr. Ken Farmer. I cannot imagine a place I would rather work or a place with a mission more worth living for! My colleagues as well deserve thanks, as they have kindly encouraged me along the way.

At the University of Cincinnati, a number of faculty have guided, supported and helped me. Dr. LihShing Leigh Wang consistently provided engaged and wise advise in course selection and dissertation development. Along with Dr. Wang, Dr. Wei Pan challenged and stretched my thinking in the classroom as well as in dissertation work. Dr. Nancy Evers and Dr. Shana Goldwyn provided helpful guidance in the early stages of dissertation work. Dr. Mary Brydon-Miller, Dr. Vicki Plano Clark and Dr. Chris Swoboda graciously and generously guided me to complete, defend and revise this dissertation. For all of these people I am grateful, and I value their contributions to my education and life.
# Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgements .............................................................................................................. v

List of Figures ......................................................................................................................... viii

List of Tables .......................................................................................................................... ix

Chapter One: Introduction ...................................................................................................... 1

   Existing Transparency Initiatives ......................................................................................... 1

   Purpose .................................................................................................................................. 5

   Hypotheses and Evidential Sources ....................................................................................... 8

   Conceptual Definitions .......................................................................................................... 13

   Significance to Higher-education Accountability Research ................................................. 16

   Delimitations ......................................................................................................................... 21

   Limitations ............................................................................................................................ 21

   Assumptions .......................................................................................................................... 22

Chapter Two: Literature Review .............................................................................................. 23

   Accountability Concerns in U.S. Higher Education ............................................................... 24

   Challenges for Accountability: The Spellings Commission and Responses ..................... 29

   Transparent Accountability Initiatives .................................................................................. 40

   Moving Forward: The Third “A” ........................................................................................ 47

   Summary Comments ............................................................................................................. 53

Chapter Three: Phase One: A Multivariate Index of Institutional Health .............................. 57

   Preliminary Methodological Notes ....................................................................................... 57

   Methods ................................................................................................................................. 58
Results.................................................................................................................................................. 71

Chapter Four: Phase Two: An Online Accountability Template ......................................................... 88

Methods................................................................................................................................................ 88

Results.................................................................................................................................................. 96

Chapter Five: Phase Three: Stakeholder Evaluation of Institutional Profiles ................................. 102

Methods............................................................................................................................................... 102

Results.................................................................................................................................................. 107

Chapter Six: Discussion......................................................................................................................... 119

Summary of Major Findings.................................................................................................................. 119

Limitations......................................................................................................................................... 126

Implications and Opportunities........................................................................................................... 127

Conclusion.......................................................................................................................................... 129

Appendix A: Screenshots for Survey of Expert Evaluators for Phase One.............................. 152

Appendix B: ABHE Draft Rubrics ...................................................................................................... 167

Appendix C: Screenshots of Items for Survey of Stakeholders for Phase Two ................. 175

Appendix D: Screenshots of Sample Items for Survey of Stakeholders for Phase Three .... 180

References.......................................................................................................................................... 132
List of Figures

Figure 1. Overview of the entire project, by phase................................................................. 9

Figure 2. An example application of Metz’s (2011) normative framework. .......................... 25

Figure 3. Metz’s (2011) framework applied to this study’s approach to accountability. .......... 54

Figure 4. Metric weights grouped by component. ..................................................................... 78

Figure 6. HLM2 model structure. ................................................................................................ 107

Figure 7. Accuracy by stakeholder group, with bubble size corresponding to respondent count.................................................................................................................................................. 109

Figure 8. Percentage correct across groups, with bubbles sized by respondent count .......... 109

Figure 9. Multivariate index of institutional health showing category ranges on a unit scale. .. 121

Figure 10. Metric ordering for Stakeholders (Phase Two) and ABHE evaluators (Phase One), with higher number corresponding to greater importance.......................................................... 125
List of Tables

Table 1: Comparison of Traits for Accountability Initiatives ......................................................... 47
Table 2: Peer Evaluator Demographic Percentages with Comparative Information from the ABHE Evaulator Pool ............................................................................................................. 61
Table 3: Peer Evaluator Demographic Variation ............................................................................. 62
Table 4: Specific Metrics Included in the Index-development Survey by Category ......................... 62
Table 5: Data Structure for Cut-Point Determination ....................................................................... 68
Table 6: Missing Data for Metric Importance Ratings by Category .................................................. 72
Table 7: Missing Data for Metric Cutpoint Ratings by Category ..................................................... 73
Table 8: Count of Expert Evaluators that Support Metric Inclusion by Metric ............................... 74
Table 9: Component Matrix for Multivariate Index with Varimax Rotation .................................... 77
Table 10: Metric Weights by Component ....................................................................................... 79
Table 11: Hypothesized and Empirical Ordering of Metrics by Index Weight ................................. 80
Table 12: Cutpoints on Included Metrics with Standardized Scores ............................................. 82
Table 13: Results of the Hypothesized vs. Empirical Cutpoints ANOVA ....................................... 84
Table 14: Specific Metrics Included in the Stakeholder-importance Survey by Category ............... 91
Table 15: Institutional Characteristics for Participating and All ABHE Institutions ....................... 93
Table 16: Phase Two Respondent Count and Percentage by Missing Data Categories ................ 93
Table 17: Phase Two Respondent Demographic Percentages ......................................................... 94
Table 18: Means, Standard Deviations, Medians and Interquartile Ranges for Importance Scores for Metrics on a 7-point Scale ..................................................................................... 97
Table 19: Demographic Percentages (n) for Removed Outliers and Retained Respondents .......... 99
Table 20: Skewness and Kurtosis before and after Square-root Transformation .............................. 100
Table 21: Multivariate Analysis of Variance of Combined Dependent Variables by Respondent Role .............................................................................................................. 101

Table 22: Phase Three Missing Data Categories with Respondent Count and Percentage ........ 106

Table 23: Phase Three Respondent Demographic Percentages ................................................... 106

Table 24: Percentage of Correct Responses by Metric (All Stakeholder Categories) ............... 108

Table 25: Percentage of Respondents Correctly Answering Template-understanding Questions
 by Metric and by Stakeholder Category .............................................................................. 111

Table 26: Stakeholder Group Means for Average Weighted Accuracy ..................................... 113

Table 27: Stakeholder Group Means for Average Satisfaction with Metrics ................................. 113

Table 28: Results of the Average Accuracy ANOVA ..................................................................... 115

Table 29: ANCOVA Results for Average Global Satisfaction by Stakeholder Group Controlled
 for Average Accuracy ........................................................................................................ 115

Table 30: Stakeholder Group Means on Average Global Satisfaction ........................................ 117

Table 31: Results of the Average Global Satisfaction ANOVA .................................................. 117
Chapter One: Introduction

How can colleges hold themselves transparently accountable to stakeholders? In a time of increased accountability and transparency pressures throughout U.S. culture (Thor, 2009), higher education finds itself subject to increased scrutiny from a sometimes-dubious public and its government (Adler-Kassner & Harrington, 2010; Ewell, 2011; Malandra, 2008; Metz, 2011). As world economic worries add their own pressures to higher education, demands increase on higher education to give a clear account of how it spends its funding, as well as what it gives students in return for their invested time and money. And while past accountability initiatives have come and gone (Ewell, 2008), the added transparency emphasis in the post-Spellings Commission climate, combined with the larger milieu of open accountability, has spawned several new initiatives within higher education – most notably the Voluntary System of Accountability, the Voluntary Framework for Accountability and Transparency by Design, all discussed in some detail below. None of the existing initiatives has documented any research into stakeholders’ perceived information needs. Additionally absent is research into how stakeholders can best be satisfied – or, turned around to emphasize the institutional role, how higher-education institutions can most satisfactorily provide “transparent accountability” for their work.

Existing Transparency Initiatives

Arguably, a higher-education transparency initiative could be as small as a single institution’s disclosing some data or as large as the federal Integrated Postsecondary Education Data System (IPEDS), which collects a wide range of institutions’ descriptive data. But in the context of higher-education accountability in the aftermath of the 2006 Spellings Commission’s report (U.S. Department of Education, 2006), discussed at length below, transparency initiatives
are expected to have three basic components: wide accessibility (most probably online systems); broad institutional coverage, including student experience and learning rather than only institutional traits such as headcount and tuition; and some degree of interinstitutional comparability. There have been three systems (two still exist) fitting this description, and each of them focuses on a particular sector of U.S. higher education: the Voluntary System of Accountability, Transparency by Design (now defunct) and the Voluntary Framework of Accountability. Each is described briefly below. While these systems respond to the Spellings Commission’s concerns, the systems do not sufficiently take into account other important components of higher-education accountability – especially accreditation bodies and processes and stakeholder perspectives.

**Voluntary System of Accountability.** Jointly overseen by the Association of Public and Land-grant Universities (APLU) and the American Association of State Colleges and Universities (AASCU), the Voluntary System of Accountability (VSA) includes state universities and colleges (M.A. Miller, 2008a). Its 277 member institutions (as of January 2015) supply information to fit a standard template called the College Portrait, and portraits are posted online. Coverage includes institutional demographics, cost-of-attendance information, student-learning data as measured by one of three standardized tests and student-experience information as measured by one of four standardized surveys (M.A. Miller, 2008a; Voluntary System of Accountability, 2008).

VSA used three standardized tests to measure student learning by applying a value-added methodology to sampled freshman and senior scores (Liu, 2010). The clear intention was to create comparable results, albeit turned into five categories (well above expectation, above expectation, at expectation, below expectation, well below expectation) instead of specific scores
(Voluntary System of Accountability, 2008). This system involved comparing value-added measures across different tests: institutions choose one of the three instruments, but value-added measures are compared for institutions regardless of the test used. In November 2011 the VSA board determined that “more work” is needed in the student-learning data area (Keller, Kuh, King, Provezis, & Phillippe, 2011); unfortunately, no greater detail was provided as to what this work will involve, or why pilot-program results were unsatisfactory. In March of 2012 the National Institute for Learning Outcomes Assessment (NILOA) issued a report on VSA’s pilot study related to student-learning data (Jankowski et al., 2012), finding that “the standardized tests of student learning originally approved for inclusion in the pilot lack credibility and acceptance within a broad sweep of the higher education community which, in turn, serves to undermine institutional participation” (p. 3).

**Transparency by Design.** Focusing on institutions that specialize in adult and online/hybrid education, Transparency by Design (TbD) was an initiative administered by the WICHE Cooperative for Educational Technologies (WCET), itself a division of the Western Interstate Commission for Higher Education (WICHE) (“New website offers transparency to adult learners returning to college,” 2009; WICHE Cooperative for Educational Technologies, n.d.). Similar to the other programs considered here, TbD included college profiles, College Choices for Adults, that included descriptive data as well as standardized information about student experiences and learning, although few institutions posted student-learning results.

What most distinguished TbD is its ability to “compare” institutions in the side-by-side way envisioned by some members of the Spellings Commission (C. Miller, n.d.). System users could select up to three institutions and view their profiles in parallel, enabling easier comparisons (although differences among institutional profiles limited usefulness). This seemed
to respond to the Spellings Commission’s interest in a database allowing such direct institutional comparisons (Commission chair Charles Miller once suggested something like a website allowing one to shop for a vehicle) (C. Miller, n.d.; U.S. Department of Education, 2006). It should be noted that this idea drew strong criticism (Ewell, 2009; Hawkins, 2008; Spiegel, 2008): not only is this concept’s feasibility questionable, but more importantly it is not clear that it makes conceptual sense to allow every individual to create his/her own, personalized ranking of institutions. Citing lack of new institutional members, a plateau in online traffic to the site and the existence of other frameworks without uniformity, the College Choices for Adults executive committee unanimously voted in July 2012 to end the program (Robinson, 2013).

**Voluntary Framework of Accountability.** Developed by the American Association of Community Colleges, the Voluntary Framework of Accountability (VFA) provides a VSA-like system for the community-college sector and has 161 participants (as of January 2015) (American Association of Community Colleges, 2011b). During its development, VFA developers had planned to utilize a transparency framework developed by the National Institute for Learning Outcomes Assessment (NILOA), which would have taken the VFA in a slightly different direction from the VSA’s standardized portraits. The NILOA framework provides a structure, but not a standard template: the framework has six components – student learning outcomes statements, assessment plans, assessment resources, current assessment activities, evidence of student learning, use of student learning evidence – each of which should be prominently presented, clearly worded, regularly updated by institutions, which make themselves open to feedback from viewers (National Institute for Learning Outcomes Assessment, 2011b). The existing public part of the VFA does not seem connected with the NILOA framework – the NILOA framework focuses on six categories of information to be published, none of which are
included in the VFA public institutional profiles. Further, VFA does not allow for public comparisons, reserving that information for member institutions to use in benchmarking. While the VFA website reports that data should have been entered by June 30, 2014, available public information as of December 2014 seems limited to minimal demographics.¹ Members are not required to submit all data.

**Purpose**

Against this background then, the purpose for this dissertation research was to refocus accountability methods by taking into account the diversified, sectored nature of U.S. higher education. Not only does U.S. higher education include narrowly focused institutions (e.g., conservatories, art schools, nursing colleges, Bible colleges), but within comprehensive universities, programs operate with a great deal of independence from each other, often having their own program-level accreditation (e.g., CAEP for education programs, CCNE and NLNAC for nursing programs, etc.). Sectors can be identified by the existence of such programmatic accrediting bodies, which are recognized by the U.S. Department of Education. By focusing accountability within sectors, richer comparisons can be enabled, moving beyond basic general-education testing to include sector-specific outcomes or concerns.

To demonstrate how this may be done, the current study sought to develop and test an online accountability template for a specific higher-education sector, the biblical higher-education sector as identified by a specialized accrediting agency, the Association for Biblical Higher Education (ABHE). It was believed that institutions in a purpose-focused sector have more commonality than institutions in the VSA, TbD and VFA sectors: institutions of biblical higher education – traditionally called “Bible colleges,” although many are now universities – share with the broader higher-education world such typical traits as general education

¹ Not all profiles were reviewed, but none of those reviewed included any data beyond demographics.
requirements, but also share among themselves a curricular focus on Christian theology/biblical studies and a holistic approach to student development.

This study aimed to measure stakeholder satisfaction within the specific Bible-college sector: this included satisfaction with (a) an online accountability display or template and (b) specific measures of institutional health, including some measures broadly applicable in higher education as well as sector-specific measures.

Part of the transparent accountability template-development included creating a multivariable index of overall institutional health, related to ABHE accreditation processes and standards. This responded to a common theme in higher-education accountability literature, the question of accreditation’s role (Brittingham, 2008; Council for Higher Education Accreditation, 2010; Dervarics, 2007; Eaton, 2010; Eaton & Council for Higher Education Accreditation, 2003; El-Khawas, 1998; Head & Johnson, 2011; Murray, 2009). Some see accreditation as providing sufficient accountability (Brittingham, 2008; Council for Higher Education Accreditation, 2010), but others point to public lack of understanding of accreditation or other perceived problems with reliance solely on accreditation (Appleson, 2004; Ewell & Wellman, 1997; Neal, 2008). Working with accreditation personnel to develop a single index that can serve as a “thermometer” for institutional health may provide greater clarity in communicating accreditation results to public audiences. Including such an index in a public template then integrates accreditation accountability into a larger transparency system.

By testing this online template – which here means measuring stakeholder satisfaction with the template as an accountability mechanism – this study’s results advance higher education’s development of meaningful, satisfactory transparency systems.

The study had four interrelated objectives across three phases.
1. Develop an institutional-health index that combines multiple data sources and that places measured institutions into “bands” of health (At-risk, Challenged, Effective, Excelling).

2. Evaluate the content validity of the index by comparing its structure to accreditation standards, other accountability initiatives and perspectives from the literature and by review by independent experts.

3. Based on published literature, on results for Objective One and on a stakeholder survey, develop an online template to present institutional profiles, transparently presenting data for accountability to a wide range of stakeholders.

4. Measure and compare stakeholder understanding of and satisfaction with profiles developed using the template developed for Objective Three.

**Research Questions by Objective**

**Objective 1:** Develop an institutional-health index that combines multiple data sources and that places measured institutions into “bands” of health (At-risk, Challenged, Effective, Excelling).

1.1. What quantitative measures of institutional health should be included in a single multivariate index of overall institutional health?

1.2. How should these measures be weighted to compose the index?

1.3. What specific levels on each measure of institutional health should be used as “cut points” to develop bands of institutional health on the index?

**Objective 2:** Evaluate the content validity of the index by comparing its structure to accreditation standards, other accountability initiatives and perspectives from the literature and by review by independent experts.
2.1. How does the multivariate index of institutional health compare with relevant accreditation standards, existing accountability initiatives and research/opinion literature?

2.2. Does content experts’ evaluation support the validity of the index components and cutoffs?

**Objective 3:** Based on published literature, on results for Objective 1 and on a stakeholder survey, develop an online template to present institutional profiles, transparently presenting data for accountability to a wide range of stakeholders.

3.1. What should be included in an institutional transparency template based on (a) research literature, (b) expert evaluator input and (c) stakeholder input?

3.2. Do these three sources of information concur on all inclusions for an institutional transparency template?

**Objective 4:** Measure and compare stakeholder understanding of and satisfaction with profiles developed using the template developed for Objective 3.

4.1. Are there differences among stakeholder groups (administrators, staff, faculty, students, students’ parents, donors, alumni) in their ability to understand profile data?

4.2. Are stakeholders satisfied that institutional transparency profiles provide sufficient information to enable viewers to evaluate institutional health?

4.3. Are there differences among stakeholder groups in their satisfaction with profiles?

**Objectives and study structure.** These four objectives fit across three phases (see Figure 1). This dissertation’s structure aligns with the three phases, as is also indicated in Figure 1, with methods and results for each phase presented in its own chapter: Chapter 3 presents Phase One, Chapter Four presents Phase Two and Chapter Five presents Phase Three. Then,
Figure 1. Overview of the entire project, by phase.

a discussion across all phases is provided in Chapter Six. This presentation method, by phase, was used to increase continuity within the presentation of each phase.

Hypotheses and Evidential Sources

Research question 1.1. Hypothesis #1: Expert evaluators will not eliminate any presented measures from inclusion in a multivariate index. Presented measures include data from the following sources: general education standardized tests, biblical content standardized test, student satisfaction survey, student engagement survey, faculty evaluations of student learning,
accreditation-agency data collection (fulltime tuition/fees, room/board, retention rate, completion rate, percentage of fulltime faculty with doctoral degrees, total weighted factor [a financial ratio used by the U.S. Department of Education as a measure of institutional financial stability]).

Evidential sources: The survey of expert evaluators asked them whether or not each metric should be included in the index. In the transparent accountability context, it was seen as better to include too much than too little, so a simple majority of respondents was considered sufficient to include a metric.

**Research question 1.2.** Hypothesis #2: Expert evaluators will assign weights to measures in the following order\(^2\) (ordinally from most to least important): standardized measures of student learning (general education and biblical content); faculty evaluations of student learning; total weighted factor; retention rate; completion rate; fulltime tuition/fees; room/board; student engagement data; student satisfaction data; percentage of fulltime faculty with doctoral degrees.

Null Hypothesis #2: Expert evaluators will assign equal weights to the measures.

Evidential sources: Expert evaluators rated each metric’s importance on a seven-point Likert-type scale. Principal components analysis (Tabachnick & Fidell, 2007) was used to develop weights for each metric.

**Research question 1.3.** Hypothesis #3: Cutpoints derived from expert input will not correspond to a normative approach, using standard deviations from ABHE means.

Evidential sources: A standard-setting method based on the Angoff technique (Berk, 1986; Cizek, Bunch, & Koons, 2004; Reckase, 2006) was used with expert evaluators. In brief, \(^2\) This order attempted to synthesize the literature and accreditation concerns: as is discussed in chapter 2, the literature emphasizes student learning, college completion and affordability; accreditation adds concerns for institutions’ financial stability and services to students.
expert input was used to determine separate cut points for each metric included in the index; cut-point and metric-weight information was then used to develop cut points for the combined index.

**Research question 2.1.** Hypothesis #4: The developed multivariate index will match relevant accreditation standards, existing accountability initiatives and research literature in its emphases and inclusions.

Null Hypothesis #4: The developed multivariate index will not match relevant accreditation standards, existing accountability initiatives and research literature in its emphases and inclusions.

Evidential sources: Metric weights in the multivariate index showed an importance or emphasis on certain components (e.g., student learning, student satisfaction) more than others. This relative importance was compared with the same components’ prominence in accreditation standards, other accountability initiatives already developed and research/opinion literature.

**Research question 2.2.** Hypothesis #5: Content experts’ evaluation of the index will support content validity of its components and cutoffs.

Null Hypothesis #5: Content experts’ evaluation of the index will not support content validity of its components and cutoffs.

Evidential sources: A qualitative survey was administered to selected content experts. Results were analyzed to identify expert consensus.

**Research question 3.1.** Hypothesis #6: Respondents will support inclusion of all information types.

Evidential sources: Evidence to answer this research question was collected by surveying (a) expert evaluators and (b) stakeholder groups. The survey, further described below, asked respondents to rate the importance of various information types. The information types presented
on this survey come from the third source, the research literature. Some information types belong in an accountability template because of their place in the research/opinion literature. Others were found to belong in the template because of survey results. If a simple majority of any stakeholder group supported inclusion of a data type, it was included.

**Research question 3.2.** Hypothesis #7: There will be significant differences among respondent groups’ importance ratings for various information types.

Null Hypothesis #7: There will not be significant differences among respondent groups’ importance ratings for various information types.

Evidential sources: Importance was measured with a seven-point Likert-type scale on a stakeholder survey. MANOVA was used to seek for differences among the respondent groups.

**Research question 4.1.** Hypothesis #8: There will be significant differences among stakeholder groups’ accuracy in interpreting data presented in online institutional transparency profiles.

Null Hypothesis #8: There will not be significant differences among stakeholder groups’ accuracy in interpreting data presented in online institutional transparency profiles.

Evidential sources: Accuracy was measured by a survey asking respondents factual selected-response questions requiring understanding of accountability profiles. Hierarchical multivariate linear modeling with two levels (HMLM2) (Garson, 2012; Tabachnick & Fidell, 2007) was planned to check for differences among stakeholder groups’ accuracy (for this research question) and satisfaction (for research question 4.3); due to low ICC values, ANOVA and ANCOVA were used instead.
**Research question 4.2.** Hypothesis #9: On average, all stakeholder groups will be satisfied that institutional transparency profiles provide sufficient information to enable viewers to evaluate institutional health.

Alternative Hypothesis #9: Not all stakeholder groups will be satisfied that institutional transparency profiles provide sufficient information to enable viewers to evaluate institutional health.

Evidential sources: Stakeholder groups’ satisfaction was determined by seven-point Likert-type satisfaction items. If a group’s mean satisfaction score was greater than neutral (a 4 on the scale), the group was seen as satisfied.

**Research question 4.3.** Hypothesis #10: There will be significant differences among stakeholder groups’ level of satisfaction with online institutional transparency templates.

Null Hypothesis #10: There will not be significant differences among stakeholder groups’ level of satisfaction with online institutional transparency templates.

Evidential sources: See discussion for research question 4.1.

**Conceptual Definitions**

1. **Accountability** – Accountability here was defined as Graham, Lyman and Trow (1995) defined “external accountability,” the process of providing “evidence and assurance… that institutional missions are being accomplished” (p. 6). This understanding is generally shared in the literature, by supporters and critics of such accountability efforts (Kelly & Aldeman, 2010; Leveille, 2006; Peters, 1994).

2. **Accountability initiatives** – This term refers to various systematic endeavors to promote higher-education accountability in the U.S. These initiatives have some sort of organizational sponsorship and higher-education institutional participants.
3. Accreditation standards – Accrediting agencies approved by the U.S. Department of Education are required to have published standards. These standards generally appear in an accreditation handbook. In the current study, the relevant standards were those of the Association for Biblical Higher Education (ABHE).

4. Expert evaluators – In this context, expert evaluators were defined as trained peer evaluators who evaluate institutions on behalf of an accrediting agency. For ABHE, these expert evaluators are generally staff, faculty or administrators of ABHE-accredited institutions. Evaluators complete training through the accrediting agency.

5. Health of a higher-education institution – A higher-education institution’s health was defined as its ability to achieve published mission statement. This reflects an accreditation perspective, as U.S. higher-education accreditation is typically understood not as defining institutional missions, but as holding institutions accountable for achieving their own stated missions. In this context, then, a healthy institution is one capable of mission achievement, and using this definition of institutional health keeps the study’s focus on combining accreditation with the larger accountability picture.

6. Information importance – The importance of a particular kind of information (e.g., information about student learning, information about students’ experiences) to stakeholders was defined as respondents’ perceived need for that information in order to make an overall evaluation about a higher-education institution’s health.

7. Information satisfaction – Stakeholders’ satisfaction with information presented in the transparency template was defined as respondents’ perception of the information’s sufficiency to enable their evaluation of the higher-education institution’s health.
8. Institutional-health index – This index is a multivariate scale developed in this study by combining weighted scores on a variety of measures, based on input from expert evaluators.

9. Institutional transparent accountability template – The institutional transparent accountability template (or simply “template”) was defined as a Web framework, developed during this study, for displaying information about institutions. The template had designated places for specified information – for example, institution name, institutional mission statement, student-satisfaction information.

10. Institutional transparency profile – An institutional transparency template with information supplied for a specific institution becomes an institutional transparency profile. In other words, template refers to the framework, while profile refers to a display of data within that framework. This terminology allows for clarity in discussing the framework separately from specific collections of information.

11. Measures of institutional health – Measures of institutional health refers to (quantitative) data concerning an array of institutional traits, including institutional finances, student costs, student learning and student satisfaction.

12. Stakeholders – Stakeholders were defined as groups of individuals who have an interest in the health of an institution of higher education. Stakeholders may be divided into two groups: internal stakeholders are those involved in day-to-day operations, including administrators, staff, faculty and (current) students; external stakeholders are those removed from day-to-day operations, including parents of current students, alumni and donors.
13. Stakeholder understanding – Here stakeholder understanding of information was defined in terms of correct/incorrect. That is, stakeholder understanding required stakeholders to be able correctly to answer factual questions about the information. Stakeholder understanding was here considered a component of transparency. This is, admittedly, a rudimentary, basic sense of understanding and is seen not as entirely sufficient, but as a necessary precursor to more sophisticated kinds of understanding: for example, “college cost” has complexity because of differences among published price, average price for the legally defined “cohort,” average discount rate, etc., but stakeholder accuracy in reading a graph of a single tuition number may be a useful consideration prior to exploring stakeholder understanding of different cost perspectives/definitions.

14. Transparency – Transparency was defined as intelligible publication of information. This required that information be made public in ways that enabled audience understanding – that is, transparency involved the degree to which institutional accountability information was understandable for various stakeholder groups.

**Significance to Higher-education Accountability Research**

This study responded to a knowledge gap in that existing transparency initiatives have not empirically included stakeholder input to determine information needs, nor has research been conducted to measure stakeholder satisfaction with those initiatives. While the study focused on a specific sector of U.S. higher education, its results and methods can be translated into other sectors. There is some reason to believe that transparent accountability is impossible on the broadest scale (nationally) or even on smaller scales (e.g., states) if those scales fail to take into account characteristics of institutions themselves (Liu, 2011; Luke, 2011). Transparent accountability seems most achievable within mission- and trait-sensitive collections of
institutions, whether in consortia, trade associations or specialized accrediting bodies. Thus, results here likely are not simply generalizable. But generalizability was not the goal; transferability or translation was the goal.

How then are these results transferrable? First, the method of measuring stakeholder satisfaction can be transferred: similar survey items can be used in different sectors. Second, this allows for comparison of satisfaction with online accountability displays across stakeholder groups. It may be that stakeholders across multiple sectors have the same attitude toward online displays. Determining this would require replication of this study in other sectors, so this study’s results primarily enable hypotheses of stakeholder attitudes for other sectors. Third, this study’s results highlight likely stakeholder attitudes both toward general measures of student learning applicable across U.S. higher education and toward sector-specific measures. While this sector’s specific measures (e.g., a test of biblical knowledge) are not relevant to other sectors, the idea of mission-focused, sector-specific instrumentation is relevant. So a consortium or accrediting agency of, say, art schools could hypothesize that their stakeholders would have similar attitudes toward an instrument developed to measure common goals for the art-school sector.

Each objective (above) has its own importance, with the first and fourth providing the greatest contributions to the field. Indeed there is a developing field of research into higher-education accountability. NILOA is a leading actor in this, conducting or commissioning studies into how institutions already do engage in transparent accountability as well as into opportunities for improved practice. Currently accreditation seems largely semidetached from transparent accountability: while accrediting agencies expect institutions both to assess and to engage meaningfully with assessment data, the accreditation work itself, including agencies’ judgments
about institutional health, remain separate from transparent accountability. Given institutions’ significant workload to achieve and retain accredited status and the degree of transparency (to accrediting agencies) involved, it seems wasteful not to communicate accreditors’ evaluations of an institution to the institution’s stakeholders. Perhaps the key problem is accreditation’s complexity. But if accreditation “insiders” can collaborate to develop a single, multivariate index that combines several accreditation-related metrics, this index could be used to convey a simpler, overall judgment concerning the institution’s “health” to the public.

The multivariate index was intended to become a public accountability mechanism, but not a tool for ranking all ABHE institutions against each other. Ranking systems incorporate a number of institutional variables in some weighted method to arrive at a single number. But when presented to those unfamiliar with the ranking system – and sometimes the specific ranking weights may not be made public – that number can gain undue power. It can be interpreted as a holistic measure of institutional quality without sufficient attention being given to the rank’s components, which can vary widely across systems (Richards & Coddington, 2010). There is some evidence that published rankings of institutions have damaged student access and student choice, among other concerns with rankings’ potential for harm (Sanoff, Usher, Savino, & Clarke, 2007). Furthermore, publishing institution-specific rankings on this index would run counter to ABHE’s approach to accreditation and accountability, which includes keeping much accreditation-related information (such as team visit reports) private and confidential.

It is important to avoid the real or perceived dangers of rankings in order to maximize the index’s attraction and utility for ABHE and, especially, for member institutions. In order simultaneously to make the index a means of achieving some level of public accountability,

---

3 There is evidence that this may be changing. The Higher Learning Commission, largest of the regional accrediting bodies, has begun publishing action letters on its website.
“banding” was used: ranges of index values that could be given broad qualitative labels were identified, and an institution’s band location could then be publicly reported. This approach has also been called clustering, tiering or grouping and has been recommended as “more appropriate and methodologically sensible” than using numerical rankings as they “help to avoid misinterpretation of small differences in indicators” and “reduce the impact of the so-called interval effect” (Merisotis, 2002, p. 479). Expert input was used in developing these ranges as well. Separately from the metric-weighting process, but in the same survey, respondents were asked to identify, for each individual metric, differences from ABHE averages (or other benchmarks) that would serve as “cut points” for the following categories: At-risk institution, Challenged institution, Effective institution, Excelling institution. Respondent means were again used, in combination with metric weights, to identify cut points on the index, creating four “bands” of institutional status. This would allow institutions to report their overall band location without publicly disclosing a single, specific index score.

Expert input represents a critical component for success of the multivariate index. Because of the peer-review accreditation model in the U.S. and its history, institutions can resent accreditation pressures, seeing the accrediting agency as “outside” and removed from institutional realities. But the accrediting agency’s reviewers come from the member institutions. Using their input to develop the index meant that its components and weights reflect the combined wisdom of institutional personnel. Further, these individuals have experience evaluating institutions based on performance indicators as well as on-site interviews, tours, etc. That is, they have experience knowing how metrics correspond with on-ground reality and making judgments of overall institutional health based on such data. This made these individuals’ collective judgment ideal for developing a single index that accurately reflects
institutional health. Combining their judgment by survey tapped into the group consensus; while focus groups might provide richer or “thicker” data, it was logistically impossible to pool as many individuals’ judgment through that mechanism.⁴

Developing an online template for transparent accountability in itself was not new, as discussed above. But extending it to a new higher-education sector contributes to the field. Further, tying the template to a specialized accreditor suggests a way of making such templates more meaningful, by working with institutional sets that have enough commonality to enable richer, broader data inclusion.⁵ Further, working within an accreditation context made it logical to include the multivariate index discussed above. This provided an “overall health” measure as part of the template, something not included in existing transparency initiatives.

Objective Four may be the most significant contribution to research. If true accountability includes satisfying stakeholders’ information needs, then measuring stakeholder satisfaction with any accountability mechanism matters. Further, it is important to begin to distinguish among different audiences (stakeholder groups): there was no guarantee that the same accountability information will satisfy different groups, so measuring differences in groups’ satisfaction advances understanding of what adequate transparent accountability requires. Beginning such work within a specific sector does not eliminate its usefulness: some template components were common throughout higher education (e.g., student satisfaction with the college experience), and it was important to begin to identify what information (and what ways of displaying that

⁴ It is worth noting that reviewers certainly take into account intangible qualities of institutions, based on their experiences during site visits. But in writing their team report – and making accreditation recommendations – they must write in response to specific delineated standards, ABHE’s Comprehensive Integrated Standards. Team members also complete and submit to ABHE worksheets asking specific questions directing team members to review/obtain evidence relevant to the standards. Thus, while intangible evidence matters, accreditation processes depend most heavily on documented institutional performance/behavior/standing.
⁵ Liu (2011) has noted the problem in the VSA of comparing institutions that are too unlike each other.
information) satisfies stakeholders’ perceived information needs – put simply, “How can we satisfy stakeholders? What information do they want? How can we clearly convey it to them?”

**Delimitations**

This study’s participants (evaluators, stakeholders) were delimited to a specific higher-education sector, biblical higher education in the U.S. as embodied in the accredited institutions of the Association for Biblical Higher Education (ABHE), the U.S. Department of Education’s designated accrediting agency for biblical higher education. This delimited identification of expert evaluators (accreditation team members within this sector) as well as institutions and their stakeholders. Such a delimitation prevents simple generalization of results to all of U.S. higher education. But resulting information may be transferrable or translatable, in whole or in part, to other specific sectors.

The study’s third phase was further delimited by use of purposive sampling. Institutions were recruited from those affiliated with ABHE: this included institutions that were (a) accredited, (b) candidates for accreditation, (c) applicants for accreditation or (d) affiliates. Participating institutions were asked to provide contact information for their stakeholders or to distribute survey information to stakeholder themselves. All institutions used the second option.

**Limitations**

As indicated above, delimitation of this study’s scope to biblical higher education limits applicability of findings. While they are not generalizable to all of higher education, or all of U.S. higher education, individual findings here may have import for higher education and may have transferability to other specific sectors of higher education. For example, within the biblical higher-education sector it is possible to have a standardized test of biblical content knowledge: while this would not matter to state institutions, it would be possible for, say, library schools
(another specialized-accreditation sector) to identify a similar core body of knowledge and use a common instrument for increased comparability.

The question of practical implications of stakeholder satisfaction was beyond this study’s scope: measuring that would require (a) actual data about actual institutions and (b) longitudinal tracking (do donors continue to give? do staff and faculty continue their employment? do students transfer or persist? etc.). This study focused on respondents’ perceptions, not on their behavior in light of those perceptions.

Results here are most limited by the small number of respondents, particularly in Phase Three. Generalizability was not a study goal. But the small dataset for Phase Three – and to a lesser extent the limited responses in Phase One – raise significant questions about the reliability of specific results here. This does not, though, limit the methodology or its transferability.  

Assumptions

The study rests on the following assumptions: (a) experts rated metrics for inclusion in the multivariate index from a position of expert knowledge based on accreditation-team experience; (b) experts rated the metrics and, later, the institutional profiles based on their best judgment; (c) stakeholders expended sufficient effort to try to understand information in the transparency template; (d) stakeholders completed the satisfaction items honestly; (e) place and time of (online) survey administration did not affect results.
Chapter Two: Literature Review

Within the last 15 years the U.S. higher-education community has paid a great deal of attention to three issues that might be called higher education’s “3 As”: accountability, accreditation and assessment (Ewell, 2008). Frequently two of these three have been discussed together in various combinations, but it has been rare for all three to co-occur. For example, some critics of U.S. higher education have emphasized the problems with accreditation structures, seeing in them a lack of accountability (e.g., Carey, 2007; C. Miller & Carey, 2007; Neal, 2008). Other voices have called for increasing accountability through various applications of assessment (Barlow, Liparulo, & Reynolds, 2007; Basken, 2008; Dwyer, Millett, & Payne, 2006) without connecting accreditation.

Moving toward some integration of the three in this study built on research literature in accountability, which includes assessment’s role in accountability efforts. This literature review begins by considering the nature of accountability in higher education in the U.S., then links accountability to assessment. Against this more general background, the Spellings Commission’s report and responses to it are considered, with a focus on the report’s call for increased public, transparent accountability in higher education. This leads to exploration of existing accountability initiatives, comparing and contrasting their approaches. Turning toward the current study, the review moves to consideration of how to include accreditation along with assessment in promoting greater, clearer accountability. The review ends with a brief summary that identifies crucial points, arising from the literature, that inform this study.
Accountability Concerns in U.S. Higher Education

Accountability has been defined as “the obligation to report to others, to explain, to justify, to answer questions about how resources have been used, and to what effect” (Trow, 1996, p. 310). Trow argues that in higher education it functions in conjunction with two other “fundamental ways in which colleges and universities are linked to their surrounding and supporting societies,” the others being trust and the market (1996, p. 310). These statements suggest that Trow writes about what has been called external accountability, in which institutions are accountable to persons (or organizations) beyond employees (Graham et al., 1995). This basic understanding, then, provides a starting point for considering how higher education has interacted with accountability demands or pressures.

An accountability framework. Metz (2011) notes that discussions of higher-education accountability use the term accountability loosely and broadly. While Trow’s (1996) definition, above, may seem fairly clear, the remainder of Trow’s article demonstrates this looseness, as he uses the single term accountability to cover varying relationships and forces. Metz proposes a normative framework to bring better clarity. His framework is presented here to allow for clearer identification of various accountability pressures, initiatives, etc.

Metz’s “terse” statement of the normative framework is, “Agent A has a duty B to provide information to body C in manner D about the extent to which higher education agent E has fulfilled its duties to do F for the sake of G, where failure to do B or F grounds a duty H on the part of agent I to respond to A or E in manner J” (2011, p. 43). This framework may be applied by specifying the agents, duties, sakes and manners: Metz sees the framework as functioning across many different specifications, with the roles not universally defined. As an example, Figure 2 applies the framework to the relationship between a higher-education
Figure 2. An example application of Metz’s (2011) normative framework.

Institution and an accrediting agency. Separate figures could be created to apply the framework to the relationship between institutions and their students, or between institutions and faculty, etc. Each application would have different identifications for agents, duties, manners, etc.

Metz (2011) notes several points of increased clarity from framework implementation. First, it is possible to distinguish between “being accountable” – duties B and F – and “holding accountable” – duty H. It is also possible to distinguish between agents A and E: although they may often be identical, this provides a means of finer analysis as necessary (for example, agent A could be the college and agent E an individual academic department, in which case the college has a duty to report on the department’s effectiveness). The framework provides for separating
the most basic or critical duties – generally F – from derived, secondary duties – duties B and H. It allows for separation of monitoring bodies (body C) from enforcing agents (agent I). And especially, it emphasizes the responding action, comprising both the duty to respond (duty H) and the response mechanism (manner J). Fundamentally, the analysis enables clearer, more precise discussion of accountability and related issues. While it was necessary to address many of these components, the focus in this study was on duty B (providing information), manner D (the mechanism for information provision) and body C (the various stakeholder groups, as well as the accrediting agency).

**Accountability and assessment.** Assessment in higher education rose to prominence as a response to accountability pressures, as a mechanism for strengthening accountability (Ewell, 2008). Before bringing these two concepts (assessment and accountability) together in the current study, it is important to distinguish between two perspectives on assessment.

**Two assessment paradigms.** According to Ewell (2008), the so-called assessment movement began in the 1980s with the articulation of two views of higher-education assessment: two reports from that time period “argued that systematic evidence about what and how much students learn is an essential prerequisite” for improvement of undergraduate education, while a third report argued “that information about what graduating students know and can do provides the ultimate bottom line… to judge the effectiveness of public investments in higher education” (p. 8). Ewell (2008) refers to these two positions as different paradigms, the “improvement paradigm” and the “accountability paradigm” (p. 9).

**The assessment paradigms and accountability.** The improvement paradigm calls for faculty, especially, to implement assessments in their ongoing teaching practice in order to ascertain how well students are learning. In terms of Metz’s (2011) framework, the reporting and
monitoring all happens within the institution: the faculty agents inform themselves (possibly each other) and are therefore also the “body” and agent charged with responding to any detected failures. Essentially this is self-accountability; while Metz (2011) does not consider this concept in depth, he seems perhaps to doubt whether it can truly be considered accountability. Trow, writing alone and with others, considers this as internal accountability (Graham et al., 1995; Trow, 1996). Here, then, is a tension within the literature: Ewell, Trow and Graham et al. – writing in the mainstream of the higher-education accountability literature – consider improvement-focused assessment to be the more important kind, providing more valuable accountability in terms of driving positive change. But Metz, writing from a more philosophical and (arguably) rigorous accountability posture, questions whether or not such systems can even be considered “holding accountable” since they do not involve any transfer of information beyond the institution itself.

What Ewell (2008) calls the accountability paradigm for assessment has a primary focus on providing information beyond/outside the institution, although measurement is still implemented within the institution. This approach seeks to provide evidence to consumers (students, parents) and funders (donors and, especially for public institutions, policymakers) as to the quality of their “return on investment.” In other words, are students getting what they, their families and the government pay for? Accountability assessment requires summative, as opposed to only formative, measures, although formative measures may figure into such a system (Ewell, 2008; Kuh & Ewell, 2010; Trow, 1996). In terms of Metz’s (2011) framework, the university reports to some external body or bodies, who have some mechanism for response (“manner” of holding accountable). (Note that any information made public can be considered external reporting, since the information is not kept private within the institution.) There are different
manners of holding accountable: funding or authorizing government agencies have an accountability mechanism, by withdrawing funds; consumers ostensibly hold accountable by choosing to attend or not to attend an institution, and although Metz argues that “It seems linguistically a stretch to say that students who elect to go to a different university are ‘holding accountable’ the avoided university” (2011, p. 44), this seems the only mechanism consumers have and so represents the best “accountability” they can exercise.

Trow (1996) and Graham et al. (1995) align with Ewell in seeing “accountability assessment” as enabling external accountability. Trow (1996) sees (external) accountability and trust – two of his three connections between institutions and their communities – as alternatives to each other, noting that “efforts to strengthen [accountability] usually involve parallel efforts to weaken trust” (p. 311). The market then responds to this balanced or in-tension relationship: when parents/students trust a college, they may be more likely to choose it in the market (Trow, 1996, p. 315). But external accountability has a key role here, for markets “perform satisfactorily only when customers are well informed” (Graham et al., 1995, p. 16), which implies that trust is impossible without some level of information transparency, and transparency enables “accountability,” suggesting that, despite the inverse relationship Trow posits, trust and accountability can work in concert.

While Ewell (2008) provides strongly contrasting views of the purposes, applications and uses of the two assessment paradigms, he also notes that “the differences between these two opposing paradigms are exaggerated, and almost no existing assessment approach fully conforms to either of them” (p. 11; see also Kuh & Ewell, 2010). The distinction, though, proves useful in approaching the contentious issue of assessment in higher education and the accountability of higher education to its various constituencies. Different perspectives on assessment are real, and
this fundamental tension runs through much of the higher-education accountability discussion. Indeed, Banta (2007) uses the same distinction in wondering, in the words of her title, “Can assessment for accountability complement assessment for improvement?” Given current endeavors in accountability, it appears that higher education has decided the two can indeed work together.

Actual implementation of comprehensive assessment programs has been spotty. Ewell (2008) briefly traces the up-and-down history of institutional, state, federal and accrediting-agency efforts to raise assessment to prominence. While institutions have implemented assessment programs of varying rigor and substance, this has taken place under the improvement paradigm, with government efforts to impose accountability programs generally fizzling historically. As will be noted below, higher education is currently trying to find ways to implement “external accountability” or “accountability assessment” itself; this study seeks to advance that effort.

**Challenges for Accountability: The Spellings Commission and Responses**

Against this background, then, Secretary of Education Margaret Spellings in 2005 appointed the Secretary’s Commission on the Future of Higher Education to examine “access, affordability, quality, and accountability” in higher education (U.S. Department of Education, 2006, p. xiii; see also Ewell, 2008). The Commission comprised nineteen individuals, eighteen of whom signed the report. After a year of hearings, meetings, research and deliberation, the Commission released its report in September of 2006. Much of the Commission’s report addresses the important issues of access and affordability, but these are not of concern here. Indeed, the aftermath of the report seems largely to have overlooked (with minimal comment)

---

6 Commissioner David Ward, then president of the American Council on Education, did not sign the report (Selingo, 2006).
these issues in order to focus on those items of greatest controversy: academic quality and, especially, accountability.

**Commission’s accountability recommendations.** In its report the Commission refers to its findings and suggestions in terms that make clear its awareness of the controversy to be engendered, mentioning the report’s frankness, calling its goals “ambitious,” noting a lack of unanimity among commissioners, admitting that its recommendations are “bold.” Others, in summarizing the Commission’s work, have also found the report frank and bold, but have added less flattering comments. Ewell (2008) notes that the report is “especially critical of accreditation” (p. 11), and Steve Crow, former president of the North Central Association’s Higher Learning Commission (HLC, the largest of the six regional accrediting agencies), called the report an assault on peer accreditation in comments he made at the HLC’s 2007 annual meeting. The Commission (U.S. Department of Education, 2006) is indeed critical of existing systems of accreditation, seeing in them “significant shortcomings” (p. 15) and calling for accrediting agencies to “make performance outcomes… the core of their assessment as a priority over inputs or processes” (p. 25).

Perhaps most controversially, the Commission called for development of a national system that would provide comparative data on a wide range of institutional factors, especially measures of student learning and degree completion. The Commission envisioned an online system maintained by the U.S. Department of Education that would “have the capacity to accommodate diverse consumer preferences through standard and customizable searches and make it easy to obtain comparative information” (U.S. Department of Education, 2006, p. 22). In making this recommendation, though, the Commission did not seem to consider the incredible institutional diversity in U.S. higher education, instead implying that consumers would want to
compare, say, truck-driving academies and comprehensive universities in a single search engine. Additionally, the Commission stated that it “supports the development of a privacy-protected higher education information system that collects, analyzes and uses student-level data as a vital tool for accountability, policy-making, and consumer choice” (U.S. Department of Education, 2006, p. 22). A system allowing for comparability across institutions necessitates creation of identical or very similar kinds of data. Recognizing this, the Commission states, “Higher education institutions should measure student learning using quality-assessment data from instruments such as, for example, the Collegiate Learning Assessment… and the Measure of Academic Proficiency and Progress” (p. 24), also mentioning the National Survey of Student Engagement and its community-college counterpart and a few other instruments (U.S. Department of Education, 2006). Essentially, implementation of the Commission’s recommendations here would require specifying a set of standardized assessment instruments as the centerpiece.

Malandra (2008) neatly summarizes the Commission’s recommendations related to assessment, including the following: implement standardized measures of learning, collect data to enable meaningful comparisons and publicize assessment results. Malandra further notes that the Commission sees accreditation as a “‘hook’ to get much of this done,” recommending that accrediting agencies more heavily focus on measures of student learning, develop mechanisms for “comparisons among institutions on learning outcomes” and “make the accreditation process more open and accessible” (2008, p. 59).

**The Spellings Commission and No Child Left Behind.** It is worth commenting on connections between the Commission and No Child Left Behind (NCLB), the controversial 2001 law mandating standardized testing and accountability reporting for primary and secondary
public education. Clearly, NCLB and the Commission’s work are linked in many people’s thinking. Critics point to NCLB as an ominous portent for higher education in the Commission’s wake (Huot, 2007; Padro, 2007). Commission member Robert Zemsky (2007) states that Spellings’s response to the report links it to NCLB. And David Shulenberger, a leader of the Voluntary System of Accountability (VSA), credits “the palpable fear of No Child Left Behind” with affecting the structure of the VSA (M.A. Miller, 2008b) – which, interestingly, was in progress prior to the Spellings Commission’s report (Malandra, 2008), adding credence to Shulenberger’s connection of VSA to NCLB.

While the Commission’s call for standardized testing, the continuity of U.S. presidential administrations between the two projects and the emphasis on accountability all suggest a connection between the Commission’s report and NCLB, the two do not seem causally related. There does appear, though, to be a philosophical connection in their shared accountability framework. As discussed above, Ewell (2008) traces concern with higher-education accountability to the mid-1980s. Of course, the spur in the 1980s was the U.S. Department of Education’s *A Nation at Risk: The Imperative for Educational Reform*, itself a “critique of elementary and secondary education” (Ewell, 2008, p. 8) and therefore related to NCLB. In sum, NCLB and the Commission’s report are related, but not as closely or causally as some responses suggested: rather than NCLB leading to the Commission, they share common roots. The passage of time has demonstrated (or brought about) a practical gap between them, as Congressional action has in some cases prevented implementation of Commission recommendations (Malandra, 2008), as opposed to Congress’s enacting NCLB.

**Higher education’s uneven response.** Initially the Commission’s report – and especially the accountability recommendations – created a furor of reaction from the higher-education
community. There seemed to be universal agreement that the Commission’s findings and recommendations were controversial and important (e.g., Banta, 2007; Brittingham, 2008; de la Rosa, 2007; Ewell, 2008; Harney, 2007; Hersh, 2007; Lingenfelter, 2007; “Spellings Report Spells Out Future for Higher Ed in U.S.,” 2006). But higher education tended to react polemically, via journalistic pieces, rather than to respond in the research literature.\(^7\) There has been a response, evolving over several years, but it has come in the form of initiatives and programs, rather than in scholarship. This section reviews both the reactions (pro and con) and the eventual programmatic responses.

**Objections to the report.** In what seems a minority view, some criticize the Commission for not going far enough or not being specific enough. Reeves (2007) finds the report “long on generalities and lofty rhetoric” with “few specific recommendations” (p. 58), doing “little to provide a blueprint for reform” (p. 59).\(^8\) Similarly, Commission member Robert Zemsky (2007) argues that the report “settled for lamentation” (p. 3), focused on testing instead of teaching or learning and failed to do “something specific to move the agenda” (p. 7).

In contrast, most criticisms of the Commission’s work see it as going too far. Such criticism often involves one of two key tensions, a consumerism-education tension and a standardization-diversity tension.

**Consumerism.** The Commission’s report frequently uses business-world or consumerist language. In the report’s preamble the Commission speaks of “this consumer-driven environment” and states that “American higher education has become what, in the business

\(^7\) In this immediate aftermath, the only peer-reviewed, research-based responses (two of them) I can find simply place the Commission’s work in a broader context (Ewell, 2008) and provide a philosophical critique (Padro, 2007).

\(^8\) It should be noted that Reeves (2007) seems on a different page from most participants in this general discussion. He includes grade inflation, “silly courses and majors,” “weak admissions policies” and “the absence of solid breadth requirements” as “huge issues” ignored by the Commission (p. 59). This represents an extreme, and rare, published position.
world, would be called a mature enterprise: increasingly risk-averse, at times self-satisfied, and unduly expensive” (U.S. Department of Education, 2006, pp. xi–xii). While such language might be expected at points in a report that addresses college cost and access, the language also occurs in discussions of learning and accountability, particularly in connection with the idea of students as consumers (see pp. 22, 24). Two Commission issue papers authored or co-authored by chair Charles Miller use even more strongly consumerist language, with the word consumer often occurring (C. Miller, n.d.; C. Miller & Malandra, n.d.). Perhaps most shocking to a reader from the academy is Miller’s car-shopping comparison referenced earlier: “While shopping for a postsecondary institution is not exactly [italics added] the same as shopping for a car, many online shopping sites embody extensive flexibility” of the sort he envisions in a data-reporting system (C. Miller, n.d., p. 2).

Such language about higher education clearly struck a nerve in many critics. Even Commission member Robert Zemsky (2007) seems attuned to it, characterizing Charles Miller as “a private investor” while also noting that he “cares deeply about American higher education” (p. 2). Ewell (2008) notes that in seeking to act on Commission ideas Secretary Spellings saw changes as “justified on the grounds of consumer protection and choice” and saw a need for a “‘consumers’ guide,’ comparable to those readily available for automobiles and refrigerators” (p. 12). Padro (2007) sees in the Commission’s work an attempt “to formalize a paradigm shift” so that institutions “are no longer only traditional places of learning: they are part of a knowledge industry” (p. 97). The American Association of University Professors (2007) seems to concur, speaking critically of the report’s “vision of higher education as a marketplace,” noting that the report “formulates a sense of crisis in almost purely financial and economic terms” (n.p.). Perhaps most vehemently, Huot (2007), in an apparent effort to discredit nearly all of the report,
seeks to “reveal an important, hidden focus for the commission,” which he finds in the idea that the Commission “frame[s] the problem for higher education as how postsecondary education can be run as a business” (pp. 521-522). In sum, there is a clear tension, embodied both in the Commission’s membership and its report’s language, between a consumerist view of higher education and what Padro calls the “traditional” view – what one outside the academy might call an “ivory tower” view. The upshot of this tension is to arouse suspicion in academic readers, as this tension reflects a longer-running tension between the marketplace and the academy, a tension that shows itself in endorsement-deal controversies and concerns that research agendas will be thwarted or controlled by the interests of corporate donors (e.g., Gould, 2003; Hutt, 1983; Seba, 2000; Washburn, 2008; Webster, 1994). Some Commission critics seem to object primarily to the terminology and thus respond negatively to the report without much consideration of whatever merit individual Commission concerns might have. Perhaps here, then, is one reason the academy has not responded with more serious substance (i.e., research into public accountability).

**Diversity and standardization.** The other key tension deals with diversity (institutional as well as disciplinary) and questions of standardization. The Commission calls for “creation of a consumer-friendly information database” for purposes of institutional comparisons (U.S. Department of Education, 2006, p. 21). Although the Commission notes that such a system “should be designed to recognize the complexity of higher education” (U.S. Department of Education, 2006, p. 22), this emphasis on comparability seems to require standardized measures, which in turn implies the use of standardized assessment instruments (or at least instruments equated to a shared metric) – the sort of instruments mentioned by the Commission (see U.S. Department of Education, 2006, p. 23). It also implies emphasis on a small set of learning to be
assessed – that is, it is not possible to implement a system that includes standardized assessment of the total diversity in higher-education programs, so the Commission’s emphasis is on student engagement and general education. This apparent emphasis on what amounts to (proposed) mandated use of a small set of standardized tests has raised widespread concerns in the academy. The concerns include both technical and philosophical questions.

Concerning the concept of using standardized pretesting/posttesting to measure “value added,” Banta and Pike (2007) review research on such systems and conclude that “research does not support the use of standardized tests for this purpose [calculation of value-added scores]” (p. 14), noting concerns with reliability and validity in this particular application of standardized tests. Hamilton and Banta (2008) argue that learning assessment that is rooted in the learning context and that allows for richer data collection provides a better alternative to standardized methods: they provide illustrative examples of different faculty members assessing student learning/performance in various courses, using common rubric elements to enable data aggregation across a program. This provides disciplinary context to, say, critical thinking or writing – and this concern for contextual and field-specific learning as opposed to only general (basic or foundational) learning/skill has been a point of concern for some (Banta & Pike, 2007; Dwyer et al., 2006). Joining in the technical, as opposed to philosophical, criticisms of standardized testing, Ennis (2008) provides a discussion of the reliability and validity concerns attending standardized testing of a complex construct such as critical thinking, arguing that validity is critical and elusive and that internal-consistency reliability may be inappropriate for critical thinking, which can be seen as multidimensional.

A more philosophically oriented response to the issue of standardization also exists. The Commission itself notes the diversity of higher education in the U.S. This theme of diversity is
widely used as an argument against the possibility of using a few standardized tests to enable appropriate comparisons across institutions (American Association of University Professors, 2007; Brittingham, 2008; M.A. Miller, 2007; Shavelson, 2007; Shulman, 2007). Additionally, critics note that standardized instruments, in order to enable broad comparisons across institutions, necessarily focus on a few basic skills, however “core” or important such skills may be – and this misses other learning outcomes that may be highly prized by institutions (Shavelson, 2007; Shulman, 2007).

A few lonely voices note redeeming aspects of standardization, although perhaps not entirely the standardization the Commission had in mind. Graff and Birkenstein (2008) seek to make a “progressive case” in favor of standardization, criticizing the institutional-diversity argument and noting that colleges do seek to instill a common set of skills and understandings in students, however difficult that set may be to define formally. Similarly, while warning of the perils of test construction, Ennis (2007) does see value in standardized testing, but in use of a wide range of tests and without the goal of comparability. Shavelson (2007) also argues for the appropriate use of testing, but with a focus on specific programs rather than basic skills, something Banta (2007) also advocates – and this programmatic focus moves toward accommodating higher-education diversity.

The critiques discussed above highlight a critical point: an accountability system relying only on a standardized test (or even a combination of them) for evidence of student learning will not be acceptable to important stakeholder groups, most notably faculty. Further, it is unlikely that the standardized tests recommended by the Commission actually (validly, reliably) measure the learning that institutions and faculty believe they are transmitting to students – which is to say that the tests are not appropriate measures of the “value” that institutions seek to add to
students. These tests are not worthless – they do seem to measure certain foundational or minimal aspects of learning – but they are insufficient for solving the accountability challenge to the satisfaction of all stakeholder groups.

Support for the report. There are a few voices from within the academy – or close by – that support the Commission, at least in part. Hersh (2007) – perhaps a suspect voice since he is involved with the Collegiate Learning Assessment (CLA), an instrument recommended by the Commission\(^9\) – argues that the “Commission got it right,” that the tension between assessment for improvement and for accountability is only “apparent” because the two are “inextricably related” and that high-stakes tests are not a problem so long as “appropriate assessments are being used” (pp. 4-5). This industry voice is joined by others, though. Lingenfelter (2007) generally supports the Commission’s calls for reform as well as the call for a “student-level data system,” noting that while learning assessment is difficult, “we are developing the capacity to address” it (p. 19). And Welch (2008) argues that, while there are “legitimate concerns about what information will be collected, how accurate it will be, and how it will be used,” nonetheless “increased assessment will be good” for higher education since “institutional assessment is the road to excellence” (p. 55).

Ewell (2011) argues that “the press and third-party policy organizations,” while historically not seen as important “external” components of accountability discussions and structures, have joined the conversation and that their involvement should only be expected to increase (p. 27). An example of this role can be seen in supportive journalistic responses to the Commission’s work, some of them authored or coauthored by Commission members (C. Miller

\(^9\) Ewell (2007) suggests that the Commission’s positive evaluation of several standardized instruments was due to “an aggressive media campaign waged primarily by the developers of the CLA” (p. 12). This argument seems a bit odd, since Ewell called the CLA a “particularly promising approach” and “the best current example of an ‘authentic’ performance-based approach” in testimony to the Commission (2006).
& Carey, 2007; Vedder, 2008). One person in particular, Education Sector analyst and now policy director Kevin Carey, became a loud voice supporting Commission recommendations and similar reform measures (2008; 2009). Indeed, Carey continues advocating for pieces of the Commission’s thinking, such as an online system to match students with institutions (2011).

Programmatic responses. Despite the broad agreement—encompassing both supporters and critics of the Commission—that the Commission’s work required response, there remains a lack of research into transparent accountability. As Ewell (2011) notes, there was a statement from the Association of American Colleges and Universities (AAC&U) and from the Council for Higher Education Accreditation (CHEA) that “adopt[ed] a more proactive stance with respect to accountability” (p. 24; see also “CHEA and AAC&U Release Accountability Statement,” 2008). What responses there are have seemed to focus on various transparency efforts, rather than research into transparency itself, into stakeholder satisfaction.

On its face, this may seem reasonable: as Ewell (2008) summarizes, the Senate’s reauthorization of the Higher Education Act blocked the Department of Education’s attempts to use negotiated rule-making to implement the Commission’s accountability and transparency (i.e., reporting) recommendations. In this case, then, one might think the academy to be justified in ignoring the Commission’s specific suggestions. But involved parties and scholars of these issues seem united in their call to the academy not to ignore the report (Banta, 2007; Brittingham, 2008; Ewell, 2008; Lingenfelter, 2007), and Ewell (2011) notes growing attention from more popular (and perhaps populist) sources.

This increased attention comes in a time when hunger for accountability transcends education: it represents a social phenomenon. Barack Obama’s 2008 presidential campaign was noted for its use of social media to promote “transparency,” allowing supporters to post even
negative comments to social-network sites without removing or censoring them (Conan, 2009; see also Thor, 2009). Similarly, consumers increasingly rely on “crowd-sourced” reviews of products, such as customer reviews on Amazon.com and similar online shopping sites. Indeed, such reviews have received news attention, particularly when an author fought back against negative reviews (Lyall, 2004) and when Amazon.com was found to be charging for better-placed reviews (Italie, 1999). Thor (2009) notes that in the accountability context transparency is in demand across fields, including “medicine, high tech, real estate, fiscal policy and just about every other industry or ideology” (p. 4). In short, U.S. consumers are accustomed to having ready access to “consumer information,” preferably from multiple sources. This trend seems unlikely to change – a recent Pew study (Taylor et al., 2011) found a majority of respondents believe U.S. college education is too expensive and found respondents divided on college’s purposes – and higher education should respond proactively (e.g., Ewell, 2011; Schneider, 2009; Wineburg, 2006). Writing about accountability and assessment from the disciplinary perspective of written communication (composition), Adler-Kassner and Harrington (2010) note that “economic and social pressures that have nurtured the accountability frame continue to grow” (p. 93; see also Zumeta, 2011). Indeed, seen in this perspective, the accountability tension can be seen as playing out tensions between competing program-evaluation paradigms, expertise- and consumer-oriented evaluation systems (Fitzpatrick, Sanders, & Worthen, 2011).

**Transparent Accountability Initiatives**

While higher education\(^\text{10}\) has not responded in research literature to these phenomena and pressures, it has not remained idle. Rather, several transparency/accountability initiatives have arisen, which were described in summary in Chapter One. Here, those programs are discussed in

---

\(^{10}\) As is at least suggested by the review, above, of objections to the report, there is a difference between individual responses, particularly by higher-education faculty, and systematic responses by institutions or organizations of institutions. Here by “higher education,” I mean systems and organizations, rather than individual actors.
more depth. While not an accountability program for specific institutions, the National Institution for Learning Outcomes Assessment (NILOA) has made important contributions to this developing field and is also presented.

**The Voluntary System of Accountability (VSA).** The oldest of the transparency programs considered here is the Voluntary System of Accountability (VSA), created under the auspices of both the Association of Public and Land-grant Universities (APLU, formerly the National Association of State Universities and Land-Grant Colleges) and the American Association of State Colleges and Universities (AASCU). VSA is not properly a response to the Spellings Commission, as the VSA project was underway before the Commission’s report (Malandra, 2008; M.A. Miller, 2008b). But Hawthorne (2008) notes that the VSA’s development was influenced by interaction between APLU/AASCU and the Commission.

VSA uses a system called the College Portrait, a framework participating institutions use to report a wide range of data in three categories: consumer information, including student progress and institutional costs; student experiences and perceptions, reporting results from instruments such as the National Survey of Student Engagement (NSSE); and student learning measures (although this remains a sticking point). Ostensibly this enables comparisons among participating institutions (Voluntary System of Accountability, 2008). The system allows users to compare two institutions on several characteristics (cost, completion rate, etc.).

Clearly, the VSA does not create the sort of government-mandated, search-engine system envisioned by the Commission. Instead, it allows institutions to work together to develop ways to measure themselves, creating a “safer” environment for accountability (M.A. Miller, 2008b). But Hawthorne (2008) notes that “participation in something like the VSA will quickly become expected and, de facto, an essential practice” (p. 24), something already realized in Ohio, where
participation was essentially mandated by inclusion in the state’s strategic plan for higher education (Ohio Board of Regents, 2008; Sims, 2008).

Of particular importance here is College Portrait content and its development. According to the VSA website, the portrait’s “data elements… were identified and evaluated based on input from student/family focus groups, feedback from the higher education community, and research on higher education” (Voluntary System of Accountability, 2008). This seems, then, to include input from external stakeholders, students and families. But none of the research released on the website details this process, summarizes its results, etc. Speaking about the program, current VSA executive director Christine Keller, of APLU, characterized the VSA as developed by eighty “higher-education leaders,” with no mention of research into external stakeholders’ information needs or evaluations of the College Portrait system (Keller et al., 2011). In the same presentation, Keller announced that the VSA oversight board had met the previous day and decided that the recently concluded pilot program for assessing student learning outcomes – which used three standardized tests and was subject to significant criticism (American Association of University Professors, 2007; Banta & Pike, 2007; Brittingham, 2008; Ennis, 2008; Hamilton & Banta, 2008; M.A. Miller, 2007; Shavelson, 2007; Shulman, 2007) – was insufficient, that approaches to learning outcomes need to be expanded, with additional data types included. This opens the door to possible connection to NILOA’s framework (Keller et al., 2011). A later NILOA report on the College Portraits and the VSA involved internal stakeholders, but does not appear to have engaged broadly with external stakeholders (Jankowski et al., 2012).

The CIC/CLA Consortium. As with the VSA, the Council of Independent Colleges Collegiate Learning Assessment (CIC/CLA) Consortium (which had its final annual meeting in
2011) predated the Commission’s report. The name denotes the system’s reliance on the Collegiate Learning Assessment, a controversial instrument intended to bring “authenticity” to standardized testing (Banta, 2008; Basken, 2008; Benjamin & Clum, 2003; Borden & Young, 2008; Haertel, 1999; Klein, Benjamin, Shavelson, & Bolus, 2007; Linn, Baker, & Dunbar, 1991; Liu, 2011; Shermis, 2008). This program followed and supplemented an earlier CIC endeavor to encourage member institutions to use the National Survey of Student Engagement (NSSE), another instrument praised by the Commission.

Unlike the VSA, the CIC/CLA Consortium did not aim to create publicly available accountability reports; rather, the focus was on assessment for institutional improvement, with the addition of collaboration meetings among institutional participants to share good practices and discuss test results and interpretations (Ekman & Pelletier, 2008). The program has now concluded, ending with a final report noting diversity of institutional experiences and identifying trends and common issues. The focus remained on internal, institutional use of assessment results to drive improvement, rather than on specific accountability endeavors (Paris, 2011). Of all programs reviewed, the CIC/CLA Consortium is the least responsive to the Spellings Commission, arguing instead for institutional autonomy with added cooperation and experience-sharing among institutions. This then can lead to broad, generalizing reports of institutional success and change – but not, of course, to institution-specific accountability or comparability, or even benchmarking.

**Transparency by Design (TbD).** With roots in the Presidents’ Forum, which began in 2004, TbD also predates the Spellings Commission’s report, but its present online form (College Choices for Adults, CCFA) was developed well after the report (WICHE Cooperative for Educational Technologies, n.d.). Similar to VSA, CCFA included institutional data, student-
satisfaction data, student-engagement data and student-learning data from standardized tests ("New website offers transparency to adult learners returning to college," 2009), although available information and instrumentation differed from institution to institution (Thor, 2009). Information was provided at the program level, something different from other transparency initiatives (Thor, 2009; "Trusted Information on Returning to College," 2010), although comparative information was largely qualitative and student-learning information not necessarily comparable (i.e., institutions may use different instruments, have different outcomes, etc.). Similar to all other programs surveyed, TbD/CCFA tended to focus on a specific sector of higher education, as could be seen from its online roster of participating institutions, as well as TbD’s affiliation with WICHE. The majority of TbD institutions were private, many of them for-profit institutions. Across all TbD institutions, public and private, there was a clear emphasis on online education, and many institutions seemed to specialize in adult education.

TbD probably came the closest to realizing the Spellings Commission’s vision of a system for comparing colleges. As noted in discussions of TbD ("New website offers transparency to adult learners returning to college," 2009, “Trusted Information on Returning to College,” 2010; Thor, 2009) and quickly discernible on the system’s website, comparisons among institutions were encouraged. Users could checkmark participating institutions in order to produce side-by-side comparisons of comparable institutional data (e.g., NSSE results, headcount, etc.). Thor (2009), president of TbD member Rio Salado College, emphasized comparisons as driving positive institutional self-reflection. As noted above, the CCFA system ended in 2013.

The National Institute for Learning Outcomes Assessment (NILOA). In 2008 assessment leaders George Kuh and Stan Ikenberry formed NILOA, based at their respective
state universities (Indiana and Illinois, now permanently located at the University of Illinois) (Hutchings & Kuh, 2011; National Institute for Learning Outcomes Assessment, 2011a). In particular, NILOA has sought to address issues of transparency in assessment of student learning, developing a “transparency framework” (National Institute for Learning Outcomes Assessment, 2011b). While it was not developed with the intention that institutions would simply adopt it entirely, the framework has been adopted by numerous institutions, something NILOA supports and enables (Keller et al., 2011). The framework was developed following two NILOA studies of nearly one thousand institutional websites and their inclusion of assessment information related to student learning (Jankowski & Makela, 2010; Keller et al., 2011). The NILOA framework presents a possible template for “standardizing” online presentations of assessment, but it was developed after most of the initiatives discussed here.

The framework, presented graphically as cyclically connected components, comprises six areas: student learning outcomes statements, assessment plans, assessment resources, current assessment activities, evidence of student learning, use of student learning evidence. For each area, NILOA provides descriptions, definitions and examples to clarify what transparency in each area looks like, and the NILOA website includes various “featured websites” from institutions using their online presence to publicize (make transparent) assessment information for greater accountability (Keller et al., 2011; National Institute for Learning Outcomes Assessment, 2011b).

The Voluntary Framework of Accountability (VFA). The latest of the transparency initiatives considered here is the VFA, sponsored by the American Association of Community Colleges (AACC). AACC vice president Kent Phillippe, also a VFA facilitator, notes that by beginning this project later, the VFA has benefited from other programs’ experiences. As noted
above, VFA was the only program adopting NILOA’s framework (Keller et al., 2011), although as noted in Chapter One VFA seems to have moved away from the NILOA framework. VFA has essentially the same purposes and rationale as the VSA, but for the community-college sector (American Association of Community Colleges, 2011b, 2011c). Phillippe notes that students are explicitly not an audience for the VFA system; rather, the focus is on providing information to the institutions themselves (i.e., benchmarking) and to external stakeholders. Reflecting this, a two-part system is used, with a secured data system for benchmarking purposes and a public piece for stakeholder accountability. The VFA is said to have been developed by “community college leaders,” with no indication of input from external stakeholders (Keller et al., 2011). Examination of steering committee and working group members provides evidence of this “insider” focus and leadership (American Association of Community Colleges, 2011a).

**Summarizing transparency initiatives.** Clearly, the Spellings Commission’s work and the discussion it launched altered our culture’s understanding of accountability for higher-education institutions. Ewell (2011) argues that it has pushed higher-education leaders toward more proactive approaches to accountability and that transparency pressures will only increase. Thor (2009) suggests that increased emphasis on transparent accountability transcends higher education and marks our entire culture. The literature in general reflects acceptance of greater accountability as simply acceptance of reality (e.g., Adler-Kassner, Linda & Harrington, Susanmarie, 2010; Malandra, 2008; Metz, 2011). Table 1 provides a comparison of the accountability initiatives on several important traits.
Moving Forward: The Third “A”

What was not, perhaps, universally foreseen in the aftermath of the Spellings Commission’s report was the response of the accrediting agencies. At the time, Brittingham (2008) emphasized the role of accrediting organizations in insisting on learning-outcomes measures. Noting that accrediting agencies have had different goals for assessment (the “improvement” paradigm) compared with the Commission’s, she also pointed to changes in regional accrediting agencies’ requirements that “have an intensified focus on student learning” (p. 35). Brittingham argued that “accreditors must… respond to the concerns of the public and

Table 1

Comparison of Traits for Accountability Initiatives

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Sector</th>
<th>Format</th>
<th>Standardized tests</th>
<th>Benchmarked data</th>
<th>Direct comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voluntary System of Accountability</td>
<td>State universities</td>
<td>Multipage online</td>
<td>Yes</td>
<td>No</td>
<td>Yes, 2 institutions</td>
</tr>
<tr>
<td>Voluntary Framework of Accountability</td>
<td>Community colleges</td>
<td>Single-screen online</td>
<td>Yes&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Not publically</td>
</tr>
<tr>
<td>CIC/CLA Consortium</td>
<td>Private liberal-arts institutions</td>
<td>No public information</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Transparency by Design</td>
<td>Online and adult education; also specific programs</td>
<td>Single Web page; can display up to three institutions</td>
<td>Yes</td>
<td>No</td>
<td>Yes, 3 institutions</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup>VFA keeps test and benchmark data private to institutions, not available to the public.
policy makers” (p. 36). This means that while institutions must “identify, define, measure, record, and evaluate student success based on measures important to the institution’s mission,” there is also a necessity “for accreditors to make judgments about that assessment” (p. 37). While the most controversial Commission recommendations were blocked by Congress, the Department of Education has moved to tighten control over higher education. Most controversial of recent actions have been the Department’s move to define the credit hour and to require distance-education providers to obtain authorization from every state in which a student resides, which brought strong responses from the accreditation community (Broad, 2010, 2011; Manning, 2010).

The tendency, then, has been to move away from direct assault on accreditation, such as through legislation, and instead to use the bureaucracy’s regulatory powers to apply pressure to and through accrediting agencies. The agencies, attempting to keep their certifying role in U.S. higher education, are applying accountability pressure to institutions (Ewell, 2011). This may be seen in recent (2009) survey results from higher-education leaders, which showed that assessment at institutions responds more to accreditation pressures than governmental pressures (Kuh & Ewell, 2010); it may also be seen in a proactive publication by the Council for Higher Education Accreditation (CHEA) seeking to articulate and defend accreditation’s value (Council for Higher Education Accreditation, 2010). Higher-education practitioners seem to value accreditation and its accountability requirements, truly seeing them as beneficial for institutional improvement (Head & Johnson, 2011; Kuh & Ewell, 2010). Increasingly, accrediting agencies are calling for more public, more transparent accountability.

In sum, the post-Spellings vision of higher-education accountability has three key emphases. First, there is an emphasis on public transparency in an intelligible fashion. Second,
there is an emphasis on assessment, especially of student learning. Third, accountability pressures are being exerted by the federal government, but working through accrediting agencies. This study, then, seeks to include accreditation in the development of public accountability, something not included in other accountability initiatives.

The Association for Biblical Higher Education (ABHE). The current study took place within the context of a specific accrediting agency, the Association for Biblical Higher Education (ABHE). Founded in 1947, ABHE has from its inception been part of a specific higher-education sector, the “Bible College movement.” Member institutions’ programs all require a curricular core in biblical/theological studies, along with general education and program-specific coursework; while Bible colleges are expanding in curricular offerings, they tend to focus on ministry-related education (Association for Biblical Higher Education, 2011b). As of 2011 ABHE included over 100 accredited, candidate, applicant or affiliated institutions enrolling over 48,000 students (Enlow, Jr., 2011).

ABHE holds recognition from the Council for Higher Education Accreditation (CHEA) and serves as the U.S. Department of Education’s recognized accrediting agency for Bible colleges. ABHE is also recognized by the International Council for Evangelical Theological Education (ICETE) (Association for Biblical Higher Education, 2011a).

ABHE seeks to become recognized as a leader in accountability within the accreditation community. The current study aligns with ABHE leaders’ desire to advance the agency, to place Bible-college accreditation “ahead of the pack” in responding to cultural, governmental and stakeholder desires for increased transparency and accountability (Enlow, Jr., 2011; also R. Bell, personal communication, February 25, 2011).
**Addressing challenges to existing transparency systems.** The transparency initiatives surveyed above link accountability and assessment, to varying degrees. But none of them includes accreditation. This first poses a challenge because, as has been discussed, accountability pressure from the U.S. Department of Education is being routed through accrediting agencies. Transparency initiatives disconnected from accreditation processes may then be less successful at addressing the accountability concerns of external stakeholders, especially government oversight bodies. Beyond this, there are additional challenges to existing transparency systems. Some of these are acknowledged in the literature or by the systems themselves. Others seem to have been overlooked.

**Commonality.** As emerged in the discussion of transparency initiatives, each tends to focus on a specific higher-education sector: VSA includes state university systems, VFA includes community colleges, CIC/CLA Consortium included private colleges and TbD included adult/online-focused institutions. But these sectors remain quite broad, in terms of being able to compare them in a meaningful fashion. In the student-learning area, Liu (2011) has noted that institutional differences may create “differential implications” for learning-outcome data (p. 7). Working through a more focused consortium or a specialized accrediting agency, such as ABHE, would allow for more meaningful, valid comparisons. This sector focus does not preclude broader comparability: it would be possible to have a transparent accountability system of online institutional profiles across multiple sectors, but there would still need to be specialized (sector-specific) information. After all, the Spellings Commission argued as if students simply choose institutions, but students also choose specific programs to prepare them for specific careers. A system that simply compares institutions on a small shared core (for instance, the general-education tests lauded by the Commission) does not help a student know about institutions’
performance in a specific program. It would be far superior to enable comparisons between, say, business programs at multiple institutions; while this was included in TbD, there was a lack of common objectives/measures/instruments across programs which limits comparability. The current study aims to move the conversation toward greater comparability: this process starts by developing ways to compare sector-specific institutions (or programs/schools within universities). This approach could be adopted by other specialized accrediting agencies. Then, institutions could link together these more focused accountability profiles into an institution-level presentation: thus a university’s overall presentation would include program-specific information, with comparability coming at the program level, not only at the institution level.

Additional measures. There is some skepticism concerning the VSA’s approach to measuring student learning purely through a single standardized test. Among others, Liu – herself involved with developing and analyzing one of the VSA tests, ETS’s Proficiency Profile – has raised questions about the value-added method used in the VSA (2010; see also Banta & Pike, 2007) as well as about the comparability of the three VSA-approved tests and issues of student motivation (2011). Despite several years of a pilot program with the standardized tests, the VSA’s oversight board has acknowledged the need for more measures without specifying what that involves (Keller et al., 2011).

Few if any voices are calling for rejection of standardized tests in these accountability structures, but there is interest in finding ways to include additional measures. In particular, there is support for use of portfolios (Banta, 2007; Basken, 2008, 2008; Eaton, 2008; Peter D. Hart Research Associates, 2008; Schneider, 2009). Adler-Kassner and Harrington (2010) note that some institutions have already developed working models in this area, such as IUPUI and Portland State University, to which could be added Washington State University, where
portfolios and other forward-thinking assessment methods (such as use of external evaluators to assess student projects) have been developed and piloted (Peterson, 2009). Related to portfolios is curriculum-embedded assessment. IUPUI faculty throughout the institution use common rubrics to evaluate students against learning objectives based on student work within a course, with data aggregated and analyzed across class levels, programs, schools, etc. (Graunke & Pike, 2011).

As indicated in the preceding discussion, there exists support for standardized measures and for contextualized, embedded measures of student learning. The current study will include both kinds of measures, with contextualized measures included in a way that allows for comparisons. This will enable measurement of stakeholder groups’ satisfaction with the different types and with a template system that incorporates both.

**Benchmarking.** Alone among the transparency initiatives surveyed, VFA explicitly includes benchmarking, although not publically (Keller et al., 2011); this should not be surprising, as the community-college sector seems to be advanced in this area (Ewell, 2011; Seybert, 2006). TbD’s side-by-side comparison system allowed a sort of benchmarking by comparing institutions, but this is not the same as providing averages from a group of related or similar institutions to use as a yardstick. Again, working within a consortium or specialized accrediting agency makes benchmarking more feasible. Data collection on common instruments would enable development of benchmarks that could then serve as a public yardstick. Within the ABHE context there is some precedent for this: a number of institutions have already participated in a benchmarking project around Noel-Levitz’s Student Satisfaction Inventory.

**Stakeholder input.** Perhaps the biggest gap in existing transparency initiatives has to do with stakeholder input. As highlighted in the discussion above, none of the initiatives has
documented research into external stakeholders’ perceived information needs, or even satisfaction with developed transparency frameworks. Yet a focus on external stakeholders was the primary motivation in the Spellings Commission’s recommendations (U.S. Department of Education, 2006), and the need for intelligible, useful information for such constituencies has been recognized in the literature (Adler-Kassner & Harrington, 2010; Ewell, 2011; Malandra, 2008; Metz, 2011; Trow, 1996). In short, a satisfactory transparent-accountability system must satisfy the information needs of various stakeholders, both internal and external, and this requires measurement. There is a need for research into both stakeholders’ perceived information needs (what do they want to know? what do they think matters?) and their satisfaction with developed accountability systems (does this satisfy them?). Working with a specific sector, such as Bible colleges, allows for more focused identification of stakeholder populations. It also addresses greater complexity, since Bible-college stakeholders are concerned with such universal issues as general education and program-specific outcomes, but also with affective student-development concerns and spiritual-development outcomes.

Summary Comments

The foregoing literature review has been broad: this was necessary because of the complex and interrelated nature of the issues involved in this study. Here the literature review is summarized into points that drove the current study.

- Metz’s (2011) framework can be specified for the current study’s accountability system (see Figure 3). This shows inclusion of multiple stakeholder groups (both as Body C and Agents I); use of online templates (Manner D); and the presumed consequences that motivate institutional improvement (Manners J).
Implied in this study was a way of combining the two assessment paradigms, improvement and accountability. If assessment information is used in an accountability system, this public accountability – and its possible consequences – motivates institutions to use that same data for improvements. Without the public accountability component, institutions have less motivation for acting on the information (the so-called “closing the loop” problem). A good transparent accountability system is not a surrender to consumerism, but an act of good faith on higher education’s part, promoting stakeholder understanding and trust. It also helps keep institutional improvement a central element of institutions’ public identity.
• The Spellings Commission’s report emphasized some concepts important for this study: availability of accountability information, ideally in an online format; comparability among institutions; intelligibility of information to stakeholders. This study did not blindly accept these emphases: first, comparability among institutions is only appropriate in some situations, when the institutions have meaningful similarity; second, whether or not accountability information is intelligible (and satisfactory) to stakeholders needed to be established empirically. The report’s controversial consumerist tone highlights the danger in developing a system that ranks institutions and presumably identifies a “best” institution: this is too simplistic for the complexity of higher education. But it may be possible to identify broader levels or “bands” of institutional performance without providing a single number that lends itself to ranking.

• Existing transparency initiatives provide both a starting point and a point of contrast.
  
  o First, no existing system seems to have included significant stakeholder input during system development, at least not in a documented way; this study surveyed stakeholders to measure their perceived information needs, which informed system structure.
  
  o The current study followed the example of VSA, TbD and VFA in including standardized tests. This accepts the need for external measures, which seem to be valued by some stakeholder groups. But it also included internal measures (faculty evaluations of student work) something valued by other stakeholder groups. While institutions are able to include such information in their profiles for VSA and TbD, this is not comparable across institutions; the current study proposed a way to achieve comparability.
Unlike VSA and TbD, this study added benchmarking, as institutional performance could be compared against ABHE-wide averages. VFA uses benchmarks, but these are not available publicly.

This study increased comparability by working with a set of institutions sharing mission-central emphases, providing for more focused standardized tests (in addition to general education). This also enabled identification of common outcome statements for use in faculty evaluations of student work.
Chapter Three: Phase One: A Multivariate Index of Institutional Health

Following general methodological discussion that applies to all study phases, this chapter focuses on an effort to develop a multivariate index of institutional health. The index was developed by surveying individuals who serve as accreditation team evaluators within the biblical higher education sector. Survey respondents were asked to identify which metrics should be included in such an index, as well as appropriate relative index weights, and then to establish cutpoints separating different levels of institutional health. Independent experts were then asked to review the index, with their responses used to evaluate index validity.

General Methodological Notes

The points discussed in this section apply to all phases and are included here as preliminary to discussion of any methods. This discussion is not repeated in Chapters Four and Five.

Statistical significance. For statistical tests, an alpha value of 0.05 was used to determine statistical significance. This conventional level was seen as a sufficiently stringent limit on false positives, given the lack of serious consequences likely to result from a Type I error in this study, but confidence intervals or precise p values are reported when possible.

Research ethics. This study was approved by the University of Cincinnati’s Institutional Review Board. All data collection was anonymous using online surveys.

Research benefits. The primary benefits from this research are to participants, both internal and external to higher-education institutions. In the absence of research into stakeholders’ information needs, institutions (and related organizations, such as the VSA and NILOA) work in a vacuum, possibly meeting stakeholder desires and possibly not providing information that will be relevant and valued. Of course, this benefit works for institutions as
well, in that they are enabled to engage in more meaningful transparency. Given the study’s accreditation context, accrediting agencies can also benefit by better understanding stakeholder views and needs. Beyond potential publication, research benefits could be promoted by presentation to professional organizations, such as the ABHE Annual Meeting.

**Researcher positionality.** In the interests of full disclosure, it is important to note that the author is himself employed at an ABHE institution; he did not participate as a respondent in the study. Further, he serves as a part of the ABHE peer evaluation corps, having completed several visits as a student services or academic evaluator. At his own institution, he serves in an administrative capacity that includes assessment and institutional research, including decisions about publication of data. His positionality increases the research’s value to him and to his institution; he hopes it also has value more broadly among institutions committed to meaningful accountability to stakeholders.

**Methods**

**Research design and rationale.** An online survey was used to measure expert opinion concerning (a) what individual metrics should be included in an overall multivariate index, (b) the importance of each metric and (c) appropriate cutpoints. An online survey enabled access to all evaluators on the ABHE roster. A survey could have been administered in person at an ABHE Annual Meeting, but evaluators are not required to attend these meetings. Also, the online survey provided clearer anonymity.

The survey presented a list of possible metrics that could be included in the multivariate index (the survey is included as Appendix A). Respondents were asked to select which metrics they believed should be included. They were also asked to rate the importance of each metric. Details of the survey design and data analysis are presented below.
**Population.** The population for Phase One comprised all peer evaluators on ABHE’s roster who have completed at least one site visit. These individuals had completed required training, given annually in conjunction with ABHE’s annual meeting. As part of site visits, they had been required to read extensive accreditation information, such as an evaluation team handbook, as well as institution-produced information. They had conducted individual and group interviews, engaged in fact-finding, deliberated with other team members, written evaluatively about the visited institution and helped make overall and specific judgments about the institution and its various attributes and functions.

Other stakeholder groups were not included in Phase One’s metric construction for two reasons. First, evaluating the metrics and suggesting cutpoints required good knowledge of higher education and the metrics; stakeholders outside higher education are unlikely to have this knowledge. Second, the index represented a distillation of some accreditation concerns and was intended to translate an “accreditation perspective” into a more widely intelligible format. This made it appropriate to use only accreditation-related individuals in metric construction.

**Participant recruitment.** ABHE’s former director of the Commission on Accreditation discussed this project with the author and expressed support and a desire to collaborate. Additionally, the current ABHE president and Commission on Accreditation director supported the study, providing letters of support.

Peer evaluators were recruited by email, with invitations extended to all peer evaluators. This aimed to capture the widest possible input into index development, increasing its “salability” to accredited institutions in the future. Demographic questions were included in the survey – personal as well as evaluation-related – so that the responding sample could be
compared with the entire roster, but due to privacy concerns, ABHE did not provide access to the roster, so demographic comparison was not possible.

**Sample.** Survey invitations were sent by ABHE staff, so the population size is not known. Thirty-five evaluators responded to the survey; three had not completed any visits, leaving thirty-two eligible participants, of which all but one contributed at least some responses. Descriptive data for this sample are presented in Tables 2 and 3. The director of the ABHE Commission on Accreditation (R. Kroll, personal communication, January 14, 2015) provided the demographic information on evaluators that is maintained; it is included in Table 2. The sample roughly matches the gender proportions in the population, but team chairs, academic evaluators and student-service evaluators are all overrepresented. ABHE does not maintain ethnicity or degree information. Based on the author’s experience attending team evaluator training and serving on teams, the age, ethnicity and completed-visit demographics do not appear to differ dramatically from the evaluator pool. The sample does not stand out as definitely inappropriate or skewed.

**Metrics included in the survey.** Survey-included metrics came from three sources: relevant literature, accreditation practice and existing transparency initiatives. Each broad metric category is discussed further below. Table 4 lists specific metrics included in the survey, by general category.

**Student learning outcomes.** Student learning outcomes (SLOs) are a sine qua non of transparent accountability: they were an emphasis in the Spellings Commission’s report (U.S. Department of Education, 2006), they are included in all transparency initiatives discussed above, and they have significant representation in the literature (Kelly & Aldeman, 2010; Kuh, 2010; Kuh & Ewell, 2010; Long, 2010; State Higher Education Executive Officers, 2008). While
there is broad agreement on SLOs’ importance, how to measure them is a “live” issue. The
Spellings Commission, emphasizing comparability, favored standardized tests (U.S. Department
of Education, 2006), which is the approach used in the transparency initiatives. The literature
includes some strong supporters of the standardized tests, perhaps even in isolation from other
measures (Klein et al., 2007; Shavelson, 2008).

Table 2

*Peer Evaluator Demographic Percentages with Comparative Information from the ABHE*

<table>
<thead>
<tr>
<th>Demographic area</th>
<th>Sample % (n)</th>
<th>ABHE % (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>21.9 (7)</td>
<td>27.7 (49)</td>
</tr>
<tr>
<td>Male</td>
<td>75.0 (24)</td>
<td>72.3 (128)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black/African American</td>
<td>3.1 (1)</td>
<td>–</td>
</tr>
<tr>
<td>White</td>
<td>93.8 (30)</td>
<td>–</td>
</tr>
<tr>
<td>Other</td>
<td>3.1 (1)</td>
<td>–</td>
</tr>
<tr>
<td>Highest degree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master's</td>
<td>40.6 (13)</td>
<td>–</td>
</tr>
<tr>
<td>Doctorate/professional</td>
<td>59.4 (19)</td>
<td>–</td>
</tr>
<tr>
<td>Experience as team chair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>34.4 (11)</td>
<td>21.5 (38)</td>
</tr>
<tr>
<td>No</td>
<td>65.6 (21)</td>
<td>78.5 (139)</td>
</tr>
<tr>
<td>Area of responsibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academics</td>
<td>62.5 (20)</td>
<td>23.2 (41)</td>
</tr>
<tr>
<td>Student services</td>
<td>18.8 (6)</td>
<td>16.9 (30)</td>
</tr>
<tr>
<td>Finances</td>
<td>3.1 (1)</td>
<td>15.3 (27)</td>
</tr>
<tr>
<td>Administration</td>
<td>15.6 (5)</td>
<td>18.1 (32)</td>
</tr>
</tbody>
</table>
Table 3

**Peer Evaluator Demographic Variation**

<table>
<thead>
<tr>
<th>Item</th>
<th>Minimum</th>
<th>25th %ile</th>
<th>Median</th>
<th>75 %ile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>35</td>
<td>50.75</td>
<td>56.5</td>
<td>62.25</td>
<td>75</td>
</tr>
<tr>
<td>Completed visits</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Years of evaluator service</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>13.75</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 4

**Specific Metrics Included in the Index-development Survey by Category**

<table>
<thead>
<tr>
<th>Broad category</th>
<th>Specific metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLOs</td>
<td>Bible Content Test freshman and senior averages</td>
</tr>
<tr>
<td></td>
<td>General Education test freshman and senior averages</td>
</tr>
<tr>
<td></td>
<td>Aggregated faculty evaluations of seniors' work using common rubrics</td>
</tr>
<tr>
<td>Student Experience</td>
<td>NSSE scores</td>
</tr>
<tr>
<td></td>
<td>SSI scores</td>
</tr>
<tr>
<td>Costs</td>
<td>Fulltime tuition and required fees</td>
</tr>
<tr>
<td></td>
<td>Room/board</td>
</tr>
<tr>
<td>Institutional Demographics</td>
<td>Retention rate</td>
</tr>
<tr>
<td></td>
<td>Completion rate</td>
</tr>
<tr>
<td></td>
<td>Percentage of faculty with doctoral degrees</td>
</tr>
<tr>
<td></td>
<td>ABHE total weighted factor (financial ratio)</td>
</tr>
</tbody>
</table>

Some work has been done in the ABHE context to identify common SLOs as well as preferred ways to assess each. A suggested rotation of instruments includes the following SLO-related tools: (a) ABHE’s Bible Content Test for biblical knowledge and (b) ETS’s Proficiency Profile or ACT’s CAAP for general education. The rotation specifies which student populations complete instruments in which years. For example, both freshmen and seniors are projected to
complete the ABHE Bible Content Test annually, while the same groups would complete the general-education test every three years. A rotation is suggested rather than simple annual administration because of assessment costs. Importantly, a goal of encouraging institutions to follow this rotation is to enable development of benchmarks through data warehousing under the accredditor’s auspices (Mort, Kavlie, & Profitt, 2011). These metrics were included on the survey for expert evaluation, but the Proficiency Profile and CAAP were evaluated as a unit, since institutions generally use one or the other.

Beyond standardized instruments, it was considered important to include a metric capturing the idea of faculty assessment of student learning. This could take various forms, such as assessment of internships or projects, of capstones or of portfolios. There are some voices in the assessment literature critical of using standardized tests alone to measure SLOs, favoring instead inclusion of other measures, especially so-called “authentic” measures such as portfolios (Adler-Kassner & Harrington, 2010; Banta, 2008; Banta & Pike, 2007; Reckase, 1997; Rhodes, 2011; Schneider, 2009). There is also some evidence that external stakeholders value such information (Peter D. Hart Research Associates, 2008). For survey purposes, it was important that this metric have enough specificity to enable expert evaluation for inclusion in the index, but there is currently no scheme widely adopted in the ABHE context. As part of the same ABHE project that identified the standardized instruments listed above, rubrics (see Appendix B) were developed for each of the common SLOs identified (Mort et al., 2011). Rather than developing something separate for this study, implementation of these rubrics was included as a possible metric: the specific measure was percentage of seniors scoring as “proficient” or “exemplary” on each rubric.
Student experience. Measures of student engagement and satisfaction are included in existing transparency initiatives. They are also supported, sometimes implicitly, by the research literature (Ewell, 2011; Hamilton & Banta, 2008) and were advocated as important by the Spellings Commission (U.S. Department of Education, 2006). Finally, existing accreditation practice commonly uses such measures.

The ABHE assessment project discussed above addressed more than SLOs, including the student experience. Specifically, the suggested instrument rotation included the National Survey of Student Engagement (NSSE) to measure student engagement and the Noel-Levitz Student Satisfaction Inventory (SSI) to measure student satisfaction (Mort et al., 2011). Beyond the benchmarking anticipated in this ABHE project, benchmarking has already been conducted for the SSI, through a partnership between ABHE and Noel-Levitz.

Costs. Along with student learning and access, cost was a key concern for the Spellings Commission. This concern is echoed in the literature and is recognized by the transparency initiatives, which include cost information. Indeed, institutions receiving federal student aid are required to include cost calculators on their websites (Adams, 2011; Fallon, 2011) and calculators are already included in institutional profiles on the Department of Education’s College Affordability and Transparency Center (http://collegecost.ed.gov/catc/Default.aspx). Current accreditation practice within ABHE also gives attention to cost. Required annual institutional reporting includes the following cost-related items: per-student tuition (published); per-student tuition (actual); required fees; room and board; and total tuition discounting. Furthermore, ABHE annually publishes information in its statistical reports aggregated across all institutions and by headcount range, allowing for easy benchmarking.
Drawing on these perspectives, two metrics were presented on the survey for possible inclusion in the index. First was fulltime undergraduate tuition and required fees; because ABHE institutions are private, residency (in or out of state) does not affect tuition. Second was the room/board total. While not all students will pay it, some ABHE institutions are highly residential, so it should be included for consideration by experts.

Institutional demographics. The term institutional demographics includes such institutional attributes as the faculty-student ratio, measures of institutional financial health, retention rates, completion rates, graduate placement, faculty credentials, etc. Collection of such indicators is common in accreditation practice, and ABHE collects a number of these in its annual report, mentioned above. Further, such indicators were addressed in the Spellings Commission’s report (U.S. Department of Education, 2006), are included in the research literature on higher-education accountability and performance (Leveille, 2006), and can be seen in state accountability efforts (e.g., Minnesota Office of Higher Education, 2009). Existing transparency initiatives include institutional demographics of various kinds, with initiatives differing from each other in their emphases.

Accreditation has been criticized for relying too heavily on metrics of the institutional-demographic type (El-Khawas, 1998; Ewell & Wellman, 1997; Ewell, Wolff, Hogan, & Banta, 1992). But such metrics do provide valuable insight into an institution’s overall health: even areas that do not directly affect student learning, such as financial stability, have far-reaching effects throughout the institution. Because this index focuses in part on providing a convenient accreditation-related measure of overall institutional health – as it were, “translating” the complexity and detail of accreditation accountability into a single indicator – this survey of experts included a number of possible metrics in this area. Some included metrics clearly related
to student learning and progress; these include the retention rate, the completion rate and the percentage of faculty with doctoral degrees. Other metrics relate to institutional finances; the survey included one of these, ABHE’s “total weighted factor”\(^{11}\) for financial health, which includes three ratios: the primary reserve ratio, which compares total expendable resources to total expenses; the equity ratio, which compares total assets to total net assets, with liabilities removed; and the net income ratio, which compares the annual change in assets to total revenue.

**Data collection.** Data were collected via an online survey system, Qualtrics (Qualtrics Labs Inc., Provo, UT), in January and February of 2013. It was estimated that the survey would take no longer than 60 minutes to complete. Respondents progressed through the following steps (see survey in Appendix A).

1. In the list of metrics, respondents selected the metrics they believed should be included in the multivariate index; they could also indicate that they lacked sufficient knowledge about a metric to suggest inclusion or exclusion.

2. Respondents rated each metric they selected on a seven-point Likert-type importance scale; again, they could indicate a lack of sufficient knowledge to rate its importance.

3. Respondents were asked to provide input to determine cut points on each metric. Each respondent was presented with all metrics sequentially, but if a respondent did not feel comfortable suggesting cut points in an area, he/she was able to leave that metric’s cut point questions blank and answer a separate question to indicate the lack of comfort with the area.

4. For each metric, benchmark or norm data were presented. Also, brief definitions of the four institutional categories (Excelling, Effective, Challenged, At-risk) were repeated for

---

\(^{11}\) It should be noted that this metric is used by the U.S. Department of Education in evaluating an institution’s capacity for administering federal student financial aid (U.S. Department of Education, 2011).
each metric, providing performance level descriptions (Cizek et al., 2004). Then, respondents were asked to identify minimum levels (cut points) for the top three index categories (Excelling, Effective and Challenged). Questions, including benchmark/norm data and category descriptions, are presented in Appendix A. This method of determining cut points is based in the standard-setting literature. In particular, it is derived from the modified Angoff method (Berk, 1986; Cizek et al., 2004; Reckase, 2006).

**Data analysis.** There were three separate analysis issues: metric inclusion, metric weighting, and cut-point determination.

*Metric inclusion.* For each metric, the proportion of respondents supporting its inclusion was determined. A metric was included in the index if a simple majority of respondents supported its inclusion.

*Metric weighting.* Respondents’ evaluations of the metrics’ importance on the seven-point Likert-type scale was used for weighting. Principal components analysis was used to determine weights (Organisation for Economic Co-operation and Development & Joint Research Centre of the European Commission, 2008).

*Cut-point determination.* Data for each cut point (minimum scores for Excelling, Effective and Challenged) for each metric were averaged across respondents. This gave a mean cut point on each metric, for each category. Table 5 shows the resulting data structure: for each metric, a minimum level for each of the listed categories was identified (the At-risk category, not shown, will simply be scores below the minimum for Challenged). These cut points on individual metrics then were combined to produce cut points for the overall index.

---

12 For all but two metrics, respondents were asked to identify absolute standards – for example, a percentage of senior test-takers whose scores must be higher than their freshman scores. But for tuition/fees and room/board respondents were asked to establish standards relative to the ABHE average. Because prices fluctuate over time and institutional sector, a norm-referenced cutpoint is more appropriate for these metrics.
Typically, normalization of individual metrics followed by additive approaches to aggregation have been used in institutional-ranking systems to derive a single overall score (Tofallis, 2012; see also Organisation for Economic Co-operation and Development & Joint Research Centre of the European Commission, 2008). Unfortunately, such methods are data-dependent in that normalization depends on the mean of the data, which changes over time and also depends on the sample or population of institutions (Tofallis, 2012). Tofallis (2012) has proposed a multiplicative approach avoiding these problems, but his approach requires all data to be on a ratio scale, which was not the case in the current study (the NSSE, SSI and total weighted factor were the data elements that cannot be measured on a ratio scale). A common additive method, the $z$-score approach, cannot be used in this case because for some metrics empirical data do not currently exist (e.g., data from NSSE and faculty use of rubrics were not yet warehoused). Instead, division by the largest value was used, which “converts the largest value on each

Table 5

*Data Structure for Cut-Point Determination*\(^{13}\)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Challenged</th>
<th>Effective</th>
<th>Excelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bible Content Test (% improved seniors)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Education test (% improved seniors)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty evaluations of seniors' work (% proficient+)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSSE benchmark scores (seniors’ mean)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSI scale scores (mean satisfaction)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuition/fees (% difference from ABHE mean)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room/board (% difference from ABHE mean)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completion rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABHE total weighted factor (financial ratio)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{13}\) Each cell will contain the minimum score for each metric-category intersection. These minimum scores will be the mean of all suggested minimum scores from expert evaluators.
criterion to unity and all others convert to a proportion of the achieved value” (Tofallis, 2012, p. 3). A slight modification from Tofallis’s method was used: rather than using the highest achieved value as the divisor, the highest possible scale value was used to avoid an institution’s ever rating over 100% on any scale.\(^{14}\) Standardization\(^{15}\) uses the following steps.

1. Cut points, for each category level on each metric, were divided by the highest possible value for that metric. This yielded three normalized scores for each metric (i.e., the boundaries between At-risk and Challenged, between Challenged and Effective, between Effective and Excelling).

2. For each metric, each of the three normalized scores was multiplied by the weight assigned to that metric.

3. The weighted cut points were added to determine index boundaries between categories (i.e., all weighted standardized scores associated with the At-risk/Challenged boundary will be added to determine the index boundary between At-risk and Challenged).

**Content validation.** Following index development, its content validity was evaluated. The following process was used to conduct a qualitative analysis of content validity (Boudreau, Gefen, & Straub, 2001; McKenzie, Wood, Kotecki, Clark, & Brey, 1999; Straub, 1989).

1. Nine experts in higher education accreditation/accountability were identified in consultation with ABHE personnel. These experts were independent from ABHE and included: a higher-education consultant who specializes in working with Christian institutions; an administrator at a Christian liberal arts university who formerly

---

\(^{14}\) In theory, tuition/fees and room/board, which are measured in terms of percentage difference from the ABHE mean, could have institutional values exceeding 100%. This is highly unlikely to occur, but were it to happen, 100% would still serve as the unity for the scale.

\(^{15}\) Tofallis (2012) calls this *normalization*. But since normalization typically refers to data transformation to produce a normal-shaped distribution, the term *standardization* is used here for clarity.
worked at an ABHE institution as a leader in distance education; an official with a regional accrediting agency; an official with an accrediting agency for Christian liberal arts institutions; two officials with NILOA; an official with VFA; a scholar at a Christian liberal arts university who has developed a standardized instrument for measuring spiritual formation; and an official with a theological-school accrediting agency. These were chosen as knowledgeable about Christian higher education, accountability/transparency or both. Six visited the survey, but only three provided responses.

2. Experts were invited to complete an online survey. The survey explained the index’s purpose and its development process.

   a. The survey asked experts to comment on the appropriateness and adequacy of each metric included in the index. Across eight items related to metrics, the average response length was 15.83 words; individual respondents averaged 21.63, 11.63 and 14.25 words across these items.

   b. The survey asked experts to comment on the appropriateness of the cutpoints identified for each metric. Across eight items related to cutpoints, the average response length was 20 words; individual respondents averaged 41.1, 12.8 and 6 words across these items.

   c. The survey asked experts to comment on the appropriateness and adequacy of the index overall (holistically). Responses to the appropriateness item averaged 8 words (individually 19, 1 and 4). Responses to the adequacy item averaged 2 words (individually 1, 1 and 4).
3. Survey data were analyzed to identify expert consensus regarding the appropriateness/adequacy of each metric, each metric’s cutpoints and the index overall.

**Results**

Results in this phase were used to develop the multivariate index. Each stage of analysis leading to index creation is presented: whether or not to include specific metrics; how each included metric should be weighted; and what cut points should be used on the overall index.

**Missing data.** The phase-one survey was the most complex in this entire study, and missing data increased as respondents moved through the survey. Two of the thirty-five respondents answered no items; one respondent provided complete demographic information but nothing else; and one respondent provided answers for all but one demographic item, then nothing else. Thus, these four respondents provided no usable data for index development.

An additional two respondents responded to demographic items, identified metrics for inclusion in an index and provided at least some importance ratings for selected metrics, but provided no responses at all on cutpoint items. Another four respondents stopped answering items part of the way through the cutpoint items.

**Metric inclusion.** Other than the four respondents who provided no information or only demographics, only one respondent had any missing data on the metric-inclusion item: that respondent did not respond for three metrics. All other respondents either (1) supported metric inclusion, (2) opposed metric inclusion or (3) reported insufficient knowledge of the metric to evaluate whether or not it should be included.

**Metric importance.** Respondents were not asked to rate a metric’s importance if they (1) opposed including it in the metric or (2) reported insufficient knowledge of the metric. Table 6
summarizes the missing data for importance ratings that this produced; note that only one case was missing for any other reason (the respondent who did not provide any rating at all for three metrics was not presented with those metrics for importance rating).

The importance ratings were used to develop metric weights to create the overall index. In order to have sufficient data to perform principal components analysis (see below), the missing data were imputed. In this case, the data were missing because respondents who opposed including the metric or reported insufficient knowledge of it were not asked to rate the metric’s importance to the index. Consequently, rather than using a more sophisticated method that would take into account response patterns from these respondents, mean imputation was used so that the average importance rating of those respondents who favored its inclusion (and did not report insufficient knowledge) was not altered. As Enders (2010) notes, mean imputation attenuated the correlations used in principle components analysis. This means that resulting weights are the minimum that would be assigned by these experts, so with more complete data a clearer picture of components and weights might be found.

Table 6

*Missing Data for Metric Importance Ratings by Category*

<table>
<thead>
<tr>
<th>Metric</th>
<th>Opposed inclusion</th>
<th>Insufficient knowledge</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABHE Bible Content Test</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>General Education test</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Faculty evaluation of seniors' work</td>
<td>7</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>NSSE</td>
<td>14</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>SSI</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tuition and fees</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Room/board</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Retention rate</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Completion rate</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>% faculty with doctoral degrees</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total weighted financial factor</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Metric cutpoints. As with inclusion and importance, four respondents provided nearly no information (none beyond demographics, if those). Two other respondents rated the importance of most metrics, but did not respond to any cutpoint questions (one reported discomfort on all cutpoints). Another three respondents stopped responding to cutpoint questions after two metrics, and a fourth stopped after three metrics. These ten respondents are included in Table 7, which shows missing cutpoint data by category. Cutpoints were calculated with pairwise deletion to maximize data use.

Metric inclusion. Metrics were included in the index if their inclusion was supported by a simple majority of ABHE evaluators responding to the survey. Table 8 includes all metrics on the survey; the nine receiving over 50% support were identified for inclusion. Two presented metrics were excluded: faculty evaluation of senior-level students’ work and the National Survey of Student Engagement (NSSE). These metrics also had the highest number of respondents reporting insufficient knowledge of the metric to evaluate its inclusion. Hypothesis 1 stated that

Table 7

Missing Data for Metric Cutpoint Ratings by Categorya

<table>
<thead>
<tr>
<th>Metric</th>
<th>Discomfort responding</th>
<th>Opposed inclusion</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Education test</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Bible Content test</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Tuition</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Room/board</td>
<td>11</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Financial factor</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Faculty degrees</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Retention</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Completion</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>SSI</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Respondents are counted in only one column, with Discomfort taking precedence over Opposed inclusion.
Table 8

*Count of Expert Evaluators that Support Metric Inclusion by Metric*

<table>
<thead>
<tr>
<th>Metric</th>
<th>Support inclusion (%)</th>
<th>Insufficient knowledge (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention rate</td>
<td>28 (90.3%)</td>
<td>1 (3.2%)</td>
</tr>
<tr>
<td>Completion rate</td>
<td>28 (90.3%)</td>
<td>1 (3.2%)</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td>26 (83.9%)</td>
<td>2 (6.5%)</td>
</tr>
<tr>
<td>Total weighted financial factor</td>
<td>23 (74.2%)</td>
<td>3 (9.7%)</td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
<td>22 (73.3%)</td>
<td>2 (6.7%)</td>
</tr>
<tr>
<td>Tuition and fees</td>
<td>21 (67.7%)</td>
<td>3 (9.7%)</td>
</tr>
<tr>
<td>Room/board</td>
<td>19 (63.3%)</td>
<td>2 (6.7%)</td>
</tr>
<tr>
<td>SSI</td>
<td>18 (58.1%)</td>
<td>2 (6.5%)</td>
</tr>
<tr>
<td>General Education test</td>
<td>17 (56.7%)</td>
<td>6 (20%)</td>
</tr>
<tr>
<td>Faculty evaluation of seniors’ work</td>
<td>15 (48.4%)</td>
<td>9 (29%)</td>
</tr>
<tr>
<td>NSSE</td>
<td>10 (32.3%)</td>
<td>7 (22.6%)</td>
</tr>
</tbody>
</table>

expert evaluators would not eliminate any presented measures from inclusion in a multivariate index. These results lead to rejection of Hypothesis 1.

Given accreditation’s large appetite for data, combined with increasing cultural and regulatory accountability pressures, it is unsurprising that expert evaluators would support inclusion of most presented measures. It appears that support for inclusion may be related to knowledge: metrics with the lowest support (General Education test scores, faculty evaluation of seniors’ work, NSSE scores) also had the highest numbers of respondents reporting insufficient knowledge to evaluate (20% or higher). NSSE was the only metric where a minority of those with knowledge supported inclusion; this may reflect NSSE’s greater expense compared with the SSI (both are in the “student experience” category). It seems that, given knowledge of the metric, the bias is toward inclusion – the more information, the better.

Differing levels of support across measures yields some surprises. The strongest support (90.3%) is for inclusion of retention and completion rates, which are also emphasized by the federal government (e.g., completion rate is included on the White House college scorecard; see
also Ewell, 2009); but these metrics are prone to practitioner criticism because of their narrow focus only on what is called the “cohort,” including only first-time, full-time, degree-seeking students who remain at a single institution. This problem with the cohort-based metric is something several organizations are working to address in the Student Achievement Measure, which is not affiliated with the government (DiSalvio, 2014; Keller, 2013). The Student Achievement Measure uses data from the National Student Clearinghouse to “track” students across participating institutions, essentially creating a “student unit record” system. There has been controversial discussion of such a system maintained by the federal government, but it has not been developed (Nelson, 2013; Stratford, 2014). Should a federal student unit record system be developed, or something like the Student Achievement Measure achieve sufficiently widespread use, more satisfactory retention and completion metrics could be used in an accountability system.

Another measure with governmental importance, the total weighted financial factor, also received high support (74.2%). Cost numbers, presumably of more importance to the public, received lower support: 67.7% for tuition and fees and 63.3% for room and board charges. This may reflect experts’ understanding that published costs often vary substantially from students’ actual cost due to financial aid and tuition-discounting models; an area for future exploration could be experts’ views on the relative utility of published cost numbers (used in this study) and cost numbers with tuition discounting taken into account.

Experts’ views on the remaining measures raise questions about their rationale, as they do not seem to present a clear pattern. The percentage of faculty with doctoral degrees, sometimes criticized as simply an input measure not necessarily related to student benefit, received the third-highest level of support (83.9%), suggesting that experts highly value faculty members and
their role. But evaluation of student learning by those same faculty (faculty evaluation of seniors’ work) received insufficient support (48.4%) to be included. Closer consideration suggests a possible explanation: faculty evaluation of seniors’ work received the highest number of “insufficient knowledge” responses (n=9); if those responses are ignored, a majority of remaining respondents (15 out of 22, 68.18%) favored inclusion.

Another apparent pair of measures, the Student Satisfaction Inventory (SSI) and the National Survey of Student Engagement (NSSE), also showed a contrast. Experts supported (58.1%) inclusion of SSI data, but not of NSSE data (32.3%). Arguably, NSSE provides better information about student learning and success, as it has been extensively investigated and connected with outputs such as student GPA (e.g., Kuh, Kinzie, Buckley, Bridges, & Hayek, 2007). This difference between expert valuation of the SSI versus the NSSE may be a question of knowledge: only 6.5% reported insufficient knowledge to evaluate the SSI for inclusion, while 22.6% reported insufficient knowledge of the NSSE.

A final contrast shows in experts’ evaluation of standardized testing. While 73.3% support inclusion of ABHE Bible Content Test data, 56.7% support inclusion of a standardized test of general-education knowledge. While the 56.7% is within the margin of error for 73.3% (which extends from 55.8% to 90.8%), it is still a striking difference between two apparently similar measures, standardized tests of student knowledge. Apart from the difference in respondents’ knowledge of the instrument already mentioned, this difference may reflect a stronger focus by experts on what distinguishes ABHE institutions, as all such institutions’ degree programs require a coursework core in Bible and theology – in other words, this test goes to the heart of what ABHE institutions have in common.
Metric weighting. Seven respondents provided a complete answer set; because of this low number, mean replacement for missing data was used as discussed above. Principal components\(^\text{16}\) extraction with varimax rotation was performed through SPSS FACTOR on the importance ratings for each of the nine included metrics. Three components were identified, each with an eigenvalue greater than 1 and together accounting for 67.29% of the variance. Table 9 shows the rotated component matrix, including the percentage of variance explained by each component; for easier interpretation, metrics are grouped by loading size.

Component 1 comprises retention and completion rates; it can be considered to measure “student persistence.” Component 2 includes room/board costs, tuition/fee costs and the total weighted financial factor, making it a measure of “costs and financial health.” Finally,

Table 9

<table>
<thead>
<tr>
<th>Metric</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention rate</td>
<td>.93</td>
<td>.10</td>
<td>.04</td>
</tr>
<tr>
<td>Completion rate</td>
<td>.91</td>
<td>.11</td>
<td>.14</td>
</tr>
<tr>
<td>Room/board</td>
<td>.06</td>
<td>.92</td>
<td>.06</td>
</tr>
<tr>
<td>Total weighted financial factor</td>
<td>.04</td>
<td>.75</td>
<td>&lt;.00</td>
</tr>
<tr>
<td>Tuition and fees</td>
<td>.49</td>
<td>.68</td>
<td>.24</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td>-.28</td>
<td>.04</td>
<td>.78</td>
</tr>
<tr>
<td>SSI</td>
<td>.15</td>
<td>.18</td>
<td>.71</td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
<td>.22</td>
<td>-.25</td>
<td>.63</td>
</tr>
<tr>
<td>General Education test</td>
<td>.16</td>
<td>.15</td>
<td>.58</td>
</tr>
<tr>
<td>Percent of variance explained</td>
<td>32.62</td>
<td>18.51</td>
<td>16.16</td>
</tr>
<tr>
<td>Percent of variance explained after rotation</td>
<td>23.29</td>
<td>22.49</td>
<td>21.51</td>
</tr>
</tbody>
</table>

Note: Factor loadings >.50 are in boldface. ABHE=Association for Biblical Higher Education; SSI=Student Satisfaction Inventory.

\(^{16}\) A component (rather than factor) method is used because there is no theory of an underlying causal factor. Instead, simple “aggregates of… variables” are sought, for which a component method is appropriate (Tabachnick & Fidell, 2007, pp. 609-610).
component 3 includes standardized tests of student knowledge, a standardized survey of student satisfaction and faculty credentials; it may best be named “student experience,” assuming that faculty credentials are seen as contributing to that experience (see Figure 4).

Weights were assigned to metrics in two steps (Berlage & Terweduwé, 1988; Nicoletti, Scarpetta, & Boylaud, 2000; Organisation for Economic Co-operation and Development & Joint Research Centre of the European Commission, 2008). First, each component was weighted by the “normalized sum of squared factor loadings” (Nicoletti et al., 2000, p. 22), which is the same as its “ratio of the sum of squared factor loadings… to the sum of the communalities” (Berlage & Terweduwé, 1988, p. 1536). For example, component 1’s weight was determined by dividing 23.29 by the sum of all three components’ values (23.29+22.49+21.51), for a component 1

![Figure 4. Metric weights grouped by component.](image-url)
weight of 0.3461. Second, each metric was weighted by the product of (1) its “normalized squared factor loading” (Nicoletti et al., 2000, p. 22) and (2) its component’s weight from step 1. Resulting weights for metrics are shown in Table 10, by component.

Null Hypothesis 2, that experts would assign equal weights to all metrics, was not supported based on the data in Table 10, since metrics were not weighted equally. It should be noted, though, that with such a small sample, standard error has not been estimated, so the null hypothesis cannot be rejected by a statistical test. Hypothesis 2 proposed a specific (ordinal) ordering of metrics. Table 11 compares hypothesized and actual (empirical) orderings. As can be seen, Hypothesis 2 was not supported, since the orderings do not align well and there were some significant mismatches, such as standardized measures of student learning (hypothesized to be most important, but ranked near the bottom by experts).

Table 10

<table>
<thead>
<tr>
<th>Metric Weights by Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics by component</td>
</tr>
<tr>
<td>Component 1</td>
</tr>
<tr>
<td>Retention rate</td>
</tr>
<tr>
<td>Completion rate</td>
</tr>
<tr>
<td>Component 2</td>
</tr>
<tr>
<td>Room/board</td>
</tr>
<tr>
<td>Total weighted financial factor</td>
</tr>
<tr>
<td>Tuition and fees</td>
</tr>
<tr>
<td>Component 3</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
</tr>
<tr>
<td>SSI</td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
</tr>
<tr>
<td>General Education test</td>
</tr>
</tbody>
</table>
Table 11

*Hypothesized and Empirical Ordering of Metrics by Index Weight*

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Order</th>
<th>Hypothesized</th>
<th>Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention rate</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Room/board</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Completion rate</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total weighted financial factor</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SSI</td>
<td>10</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Tuition and fees</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>General Education test</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Faculty evaluations of seniors' work</td>
<td>3</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>NSSE</td>
<td>9</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

The weights assigned to each metric within its component are problematic. The range of the weights is only 0.087; combined with small samples, this raises questions about whether or not these weights are meaningfully different from each other. It is still instructive to make the general observation that these experts seem to value more “objective” measures – retention and completion rates, costs, percentage of faculty with doctoral degrees – than measures less *prima facie* valid, such as standardized tests. This seems to support a common criticism of accreditation, that its processes tend to focus more on easily measured inputs than on harder-to-measure, but also more important, issues such as student growth and learning (e.g. U.S. Department of Education, 2006).

Hypothesis 4 – stating that the multivariate index’s emphases would match relevant accreditation standards, existing accountability initiatives and research literature – was evaluated using Table 11’s rank information. Hypothesis 4 was not supported by the data, although a Wilcoxon signed-rank test did not find a significant difference, $T=25$, $p>0.05$. Evaluators most heavily weighted retention and completion (component 1) as well as some costs to students.
(parts of component 2), an approach aligned with accountability pressures such as the U.S. Department of Education’s College Scorecard or the College Measures system (Bidwell, 2013). In particular, the emphasis on retention/completion aligns well with accountability initiatives and the literature. Unlike accreditation standards and research literature, as well as some accountability initiatives such as NILOA, the evaluators did not as heavily value measures of student learning, choosing not even to include faculty evaluation of student work, something supported by the literature; this may be explained by respondents’ unfamiliarity with the concept, as fewer opposed than supported its inclusion (the number reporting lack of knowledge kept it from being included).

**Cutpoint determination.** Expert respondents were also asked to identify appropriate cut points on each included metric. Averaging across responses determined minimum achievement levels for each institutional category other than At-risk (that is, for Challenged, Effective and Excelling); the At-risk category includes all scores below the Challenged minimum threshold. Table 12 shows the minima on each metric for each category – for example, on the ABHE Bible Content Test, an institution with less than 55% of students improving their scores (test/retest) would be categorized as “At risk” on that metric, while an institution with an improvement percentage between 55% and 65% would be categorized as “Challenged.”

These data were also used to determine composite cut-points on the overall index. Normalized metric cut points were weighted, based on metric weights presented above, and summed. Overall minima for each level are presented in Table 12’s last row. These values can be converted to a different scale as desired.

---

17 The survey specifically asked about faculty evaluation of seniors’ work, intending to capture the popular “capstone” approach. It may be that this specificity – rather than a more generic statement such as “faculty evaluation of student work” – caused this metric to be less understood, as discussed above, or less valued.
Experts’ relatively low expectations for student learning and satisfaction provides the most striking aspect of cutpoints (Table 12). On average, experts considered as Effective an institution where just 65% of graduates showed improvement on standardized scores of biblical content and of general education. On the question of student satisfaction, experts considered effective a performance below the national average (based on the national norms presented to the experts on the survey): the average across the scales is 5.22, while the Effective cutpoint was set at 5.03.

On cost issues (tuition/fees and room/board), experts showed some tolerance for cost differences, seeing as Excelling a range of costs within 25% of ABHE averages. But this tolerance was limited: a deviation from ABHE averages of more than 35% was seen as At risk. Similar clustering was evident in completion rates, with the cutpoints differing little from each other in standardized (z score) units (from 0.03 to 0.05) – and thus differing little from ABHE norms.

Table 12

<table>
<thead>
<tr>
<th>Cutpoints on Included Metrics with Standardized Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metric</td>
</tr>
<tr>
<td>ABHE Bible Content Test (% improved)</td>
</tr>
<tr>
<td>General Education test (% improved)</td>
</tr>
<tr>
<td>SSI scale scores (7-point scale)</td>
</tr>
<tr>
<td>Tuition/fees (% difference from ABHE mean)</td>
</tr>
<tr>
<td>Room/board (% difference from ABHE mean)</td>
</tr>
<tr>
<td>Retention rate</td>
</tr>
<tr>
<td>Completion rate</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
</tr>
<tr>
<td>Total weighted financial factor (-1 to 3)</td>
</tr>
<tr>
<td>Overall index minima</td>
</tr>
</tbody>
</table>

*Note: When possible z scores for cutpoints are given in parentheses.*
Finally, in the areas of retention rate and faculty credentials experts separated the categories further. Faculty credentials showed the largest spread, from 40% with doctoral degrees as the cutpoint between At risk and Challenged to 70% as the cutpoint between Effective and Excelling. Retention-rate differences were nearly centered on the ABHE mean for Effective (z score of -0.01), with Challenged almost one standard deviation down (z=-0.98) and Excelling 10 percentage points higher than Effective, a z score of 0.65.

Hypothesis 3 states that expert-derived cutpoints will not correspond with a normative approach. A normative approach would begin with ABHE averages, then derive cutpoints based on standard units (z scores). To test this hypothesis, data for ABHE institutions were accessed from the National Center for Education Statistics (NCES), both from the Integrated Postsecondary Education Data System (IPEDS) and from the 2004 National Study of Postsecondary Faculty. National norms for the ABHE Bible Content Test are not available publically, nor are ABHE norms for the SSI. While national norms are available for the General Education tests, that information does not include longitudinal data such as the percentage of students whose scores improve. The mean percentage of faculty with doctoral degrees is available through ABHE statistical reports, but not the deviation necessary to determine z scores. To address this, deviation data were excerpted from the 2004 National Study of Postsecondary Faculty (NSOPF:04) for faculty credentials of private, not-for-profit institutions in the specialized-theological Carnegie category. This group is likely not a perfect match with ABHE institutions, but the mean percentage of faculty with doctoral degrees is very similar, 49.5% for the NSOPF:04 group and 50% for ABHE in the most recent statistical report. Finally, while ABHE reports mean weighted financial factors, no variance is provided, making it impossible to
calculate $z$ scores for this metric; it is worth noting that both the Challenged and Effective minima are below the ABHE average factor score of 2.1.

Table 12 includes $z$ scores for metric cutpoints when possible; note that while data to calculate several $z$ scores are currently unavailable, this information could easily be generated if an accrediting agency adopted a framework and began to warehouse the information, collecting it via existing annual report mechanisms. Hypothesis 3 was tested by a $3 \times 2$ ANOVA (see Table 13). Hypothesized $z$ scores for cutpoints were -1 for Challenged, 0 for Effective and +1 for Excelling; empirical $z$ scores were as given in Table 12. The difference between hypothesized and empirical $z$ scores did not reach significance, $F(1, 14)=4.457, p=0.053$, so the null cannot be rejected. The small $p$ value, though, and the data itself show that there is not neat alignment between hypothesized and empirical cutpoints.

It seems intuitive to use deviations from ABHE averages to identify cutpoints, and while Null Hypothesis 3 was not rejected, available $z$ scores show that experts tend to prefer less variation from the mean, with all but two cutpoints falling within ±1 on a $z$ scale. Given the accountability emphasis on student learning and marketing concerns for the student experience, experts’ low expectations for these items seem particularly conservative, even dissatisfactory; it would be valuable to survey external stakeholders to determine their expectations. This cautious

Table 13

*Results of the Hypothesized vs. Empirical Cutpoints ANOVA*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>1.600</td>
<td>1</td>
<td>4.457</td>
<td>0.241</td>
<td>0.053</td>
</tr>
<tr>
<td>Error</td>
<td>5.014</td>
<td>14</td>
<td>(0.358)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6.610</td>
<td>15</td>
<td></td>
<td>(0.358)</td>
<td></td>
</tr>
</tbody>
</table>

*Note: MSE given in parentheses.*
approach by experts may come, of course, from uncertainty about the actual variation among institutions; while the survey provided averages when possible, measures of variation were not included. This study sought to use criterion references when possible, asking experts to identify cutpoints on a metric’s scale. Another approach could ask experts to identify cutpoints on all metrics in terms of deviation from the mean among ABHE institutions; this might lead to cutpoints more easily interpreted, and more easily explained to institutions and to stakeholders more broadly.

**Validation.** Experts (n=3) without direct ABHE involvement commented on (1) the appropriateness and (2) the adequacy of metrics included in the overall index. In general, responses were positive for specific metrics or categories of metrics. One respondent noted that some metrics are “not inexpensive,” and two respondents noted that some measures need clear definition: “Do you mean, for example, annual retention rates or semester to semester? And does completion rate refer to 100% of the normal time or 150% or even 200%?”

Regarding demographic-type inclusions (retention and completion rates, percentage of faculty with doctoral degrees, weighted financial factor) one respondent noted “Though most of these metrics are good, it is not exhaustive enough,” recommending addition of such measures as job placement rates and pass rates for licensure exams. These additional measures were not considered in this study for two reasons: first, many ABHE colleges do not have programs targeting licensure, giving that metric poor coverage; second, there is currently no single definition of job placement (although the U.S. Department of Education requires institutions to publish this information, institutions are free to choose how they measure it). Similarly, two respondents noted a need for additional information in the cost area, so that measures took into consideration financial aid and other tuition discounting (one respondent commented, “It only
measures published rates and not the actual cost, which is in two forms: 1) the discounted rate after scholarships, and 2) the ‘real’ cost that it takes a school to educate a student”). For simplicity, this study included only published tuition, but the respondent correctly identified this area of controversy and complexity, that is, how to measure college cost, which is discussed in the U.S. Department of Education’s draft framework for a college ratings system (2014). Prior to implementation, an accountability system would need to identify a better cost metric, which might mean developing a new composite metric.

Concerning cutpoints, one respondent suggested that all identified points seem arbitrary. The same respondent suggested modification to cutpoints on several metrics, such as making year-to-year comparisons for the same institution. Additionally, this respondent expressed concern over the usefulness of the weighted financial factor (noting that it emphasizes liquidity) and of the percentage of faculty with doctoral degrees. Another respondent challenged cutpoints on two metrics (SSI scores, percentage of faculty with doctoral degrees). The third respondent supported identified cutpoints on all metrics.

Overall, two respondents viewed the index overall as appropriate. The third respondent, though, noted that “these measures, however commendable, tend to reduce a very complex entity to an overly simplistic one.” While responses provided limited support for the index, they raised sufficient questions overall and concerning specific metrics or cutpoints as to fail to support the index’s validity. In particular, three points in expert responses were sufficiently serious to undermine validity: the credibility of cost numbers as presented; overreliance on a single measure of financial health (the financial factor); and the inadequacy of the metrics, in total, to capture institutional health. Thus, Hypothesis 5, that experts’ evaluation will support the content validity, is not supported.
Phase One Summary

While Phase One did not result in a validated multivariate index of institutional health, it did demonstrate a feasible methodology. Within a specific higher-education sector – in this case, Bible colleges, but transferable to other sectors – specialized accrediting agencies represent a body of experts. These experts can collaborate to identify accountability metrics not only applicable to higher education broadly, but also metrics for outcomes or institutional characteristics of particular importance within their sector. Further, experts can work to identify appropriate “cutpoints” for varying levels of institutional health, here defined as the ability to fulfill the institution’s mission.

This phase identified not only challenges in expert responses – gaps in individual knowledge as well as discomfort with the metric (or cutpoints) conceptually – but also particular points of complexity that complicate development of an overall index, especially measures of retention, completion, and cost. The challenges identified here point to a possible methodological improvement, to begin work on such an index with qualitative approaches, including focus groups and collaborative working groups of experts.

The next chapter turns to Phase Two. Building on and extending Phase One, Phase Two moves to public accountability systems, bringing stakeholders into the study.
Chapter Four: Phase Two: An Online Accountability Template

Phase Two focused on developing online accountability displays. An important step in this is surveying stakeholders to determine what information they want to see included in an accountability template. Chapter Four presents methods and results for Phase Two, which addressed Objective Three: Based on published literature, on results for Objective One and on a stakeholder survey, develop an online template to present institutional profiles, transparently presenting data for accountability to a wide range of stakeholders. Objective Three was comprised of two research questions: (1) What should be included in an institutional transparency template based on (a) research literature, (b) expert evaluator input and (c) stakeholder input? and (2) Do these three sources of information concur on all inclusions for an institutional transparency template?

Methods

Research design and rationale. This phase included development of an online display (template) of institutional information, intended to embody transparent accountability for ABHE institutions. An online system was chosen as most appropriate to be readily available to a wide range of stakeholders. A standardized template was chosen as the best feasible means of enabling stakeholders to compare institutions: while a system allowing stakeholders to select institutions whose data are then juxtaposed on a single screen would be more convenient for stakeholders, it also requires greater technical expertise in design and maintenance, so a standardized template presents a compromise. The research literature and Phase One’s results informed website construction, but a survey of stakeholder groups was also conducted. Seeking broad stakeholder input, this survey addressed a gap in other accountability initiatives, explicit inclusion of
stakeholder information needs. Beyond design of the website structure (template), four profiles for fictional institutions were created for use in Phase Three.

Template design involved three issues. First, how would the template be structured – how would information be organized? Second, what data would be presented? Third, what would the template look like, in design (including aesthetics)?

**Structure.** Determinations about template structure were intended to be made in light of two information sources. First, existing transparency initiatives provided examples of possible organizational structures. Second, it was expected that different stakeholder groups (e.g., students, alumni) would value different information sources, which would have led to creating templates with differing sections for stakeholder groups (e.g., a “Student” section, an “Alumni” section). This issue was eliminated from site structure decisions once Phase Two’s stakeholder survey results indicated no significant differences among the groups. Therefore, templates were structured by information type (student learning, student experience, etc.).

**Data inclusion.** Determining what specific information should be included in the template drew on four sources. First, again, existing transparency initiatives suggested some metrics as important. Second, research literature on higher-education accountability and transparency illuminated the question. Third, expert input in Phase One concerning what metrics to include in the index also had implications for what metrics to include in the template. Fourth, a wide range of stakeholder groups was surveyed to measure the importance stakeholders assign to different data sources for determining institutional health. This fourth component separates the current study from other initiatives, which did not draw on broad stakeholder input, as discussed in Chapter Two.
**Design.** The website-design tool imposes some limitations; the specific tool (discussed below) was chosen to maximize institutional usability, which means low or no cost and ease of use for personnel not necessarily familiar with HTML coding or Web design, since ABHE institutions tend not to be able to afford such specialized personnel.

**Data for templates.** While determination of specific data for template inclusion depended in part on Phase One’s results and the additional survey conducted in this phase (discussed below), it was possible to delineate categories and some likely inclusions ahead of time. The broad categories matched those used in Phase One: student learning, student experience, cost and institutional demographics. When possible, ABHE benchmarks were presented in the template (fictional in some cases, actual when possible).

**Stakeholder survey.** Beyond Phase One results, research literature and existing accountability initiatives, it was desirable to include broad stakeholder input in determining what data to include in templates. To accomplish this, internal and external stakeholders were surveyed.

**Population.** The population comprised stakeholders of ABHE institutions. Specifically, this survey targeted the following stakeholder groups, restricted to institutions’ undergraduate level: administrators, staff, faculty, current students, current students’ parents, alumni, and donors. (This includes most Phase One participants, as most ABHE evaluators are employees of member institutions.) The target population, then, included all individuals in these categories at all ABHE-affiliated institutions. At the time of the survey, there were 192 such institutions.

**Survey design.** The survey (see Appendix C) was designed and administered using Qualtrics, an online survey system; it was not piloted prior to administration but was reviewed by a faculty committee and was based on Phase One’s survey. The survey was similar to Phase
One’s survey in the metrics involved, although it added some qualitative, narrative items that were not appropriate for the Phase One index and removed the total weighted index (financial ratio), which is not easily interpreted; it also included more stakeholder groups, not only experts.

Following an informed-consent page, the survey had two basic sections.

**Respondent demographics.** To enable analysis, respondents were asked to indicate their role (administrator, staff, faculty, student, student’s parent, alumnus/alumna, donor) and their institutional affiliation (from a list of institutions participating in the study), along with basic demographic information (age, gender, ethnicity, education level).

**Information importance.** Respondents were presented with a list of possible template inclusions (see Table 14) including the elements listed above, as well as most of the metrics considered in Phase One for index inclusion. For each possible template inclusion, respondents

<table>
<thead>
<tr>
<th>Broad category</th>
<th>Specific metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Learning</td>
<td>Bible Content Test freshman and senior averages</td>
</tr>
<tr>
<td></td>
<td>General Education test freshman and senior averages</td>
</tr>
<tr>
<td></td>
<td>Aggregated faculty evaluations of seniors' work using common rubrics</td>
</tr>
<tr>
<td></td>
<td>Links to anonymous samples of student work</td>
</tr>
<tr>
<td>Student Experience</td>
<td>NSSE scores</td>
</tr>
<tr>
<td></td>
<td>SSI scores</td>
</tr>
<tr>
<td></td>
<td>Student testimonials/biographies</td>
</tr>
<tr>
<td></td>
<td>Alumni testimonials/biographies</td>
</tr>
<tr>
<td>Costs</td>
<td>Fulltime tuition and required fees</td>
</tr>
<tr>
<td></td>
<td>Room/board</td>
</tr>
<tr>
<td>Institutional Demographics</td>
<td>Retention rate</td>
</tr>
<tr>
<td></td>
<td>Completion rate</td>
</tr>
<tr>
<td></td>
<td>Percentage of faculty with doctoral degrees</td>
</tr>
</tbody>
</table>

*These metrics/information sources were not included in the Phase One survey.*
were asked to report that inclusion’s importance to them – that is, their perceived need for that information in order to make an overall evaluation about an institution’s health. This importance was measured using a 7-point Likert-type item for each aspect of the transparency template, ranging from 1=“Not important at all” to 7=“Very important.”

**Sampling, participant recruitment and data collection.** Participant recruitment began with institutional recruitment. Formal support was given by ABHE, in the form of a (digital) letter of support from ABHE president Ralph Enlow for use in recruitment.

An invitation to participate was sent (by email) to presidents, copied to chief academic officers (CAOs), of the 192 ABHE-affiliated institutions. (A follow-up email was sent a week later to institutions not responding.) This email explained the project, had the attached letter of support from ABHE and had attached instructions; seven institutions agreed to participate. Of these seven, two were not accredited (institutions may affiliate with ABHE without being accredited). Information for the other five participating institutions was retrieved from the U.S. Department of Education’s Integrated Postsecondary Education Data System (IPEDS) and comparative information was retrieved from ABHE’s 2013 statistical report ([http://www.abhe-resources.com/pdfResources/COA/2013_Statistical_Highlights.pdf](http://www.abhe-resources.com/pdfResources/COA/2013_Statistical_Highlights.pdf)). Table 15 shows the groups’ information on several characteristics (ABHE’s report does not provide a measure of variability, unfortunately). With the small sample and one of ABHE’s larger institutions, the largest mismatch between the sample and ABHE averages is in undergraduate headcount. On other characteristics, the sample aligns fairly well with ABHE averages. Still, a larger number of institutional participants would have greatly strengthened the study.

Participating institutions were asked to provide email addresses for administrators, staff, faculty, students, students’ parents, alumni and donors. It was anticipated that some institutions
Table 15

Institutional Characteristics for Participating and All ABHE Institutions

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participant M (SD)</th>
<th>ABHE M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total tuition/fees</td>
<td>$4,836,008.6 ($6,847,760.44)</td>
<td>$3,370,119.00</td>
</tr>
<tr>
<td>Total gifts</td>
<td>$1,633,013.80 ($1,076,298.54)</td>
<td>$1,967,127.00</td>
</tr>
<tr>
<td>Full-time instructional staff</td>
<td>13.20 (7.79)</td>
<td>12.70</td>
</tr>
<tr>
<td>Retention rate</td>
<td>66.60 (12.12)</td>
<td>75.10</td>
</tr>
<tr>
<td>Undergraduate annual headcount</td>
<td>904.20 (1285.02)</td>
<td>405.40</td>
</tr>
<tr>
<td>Graduation rate</td>
<td>46.00 (6.52)</td>
<td>48.00</td>
</tr>
<tr>
<td>Undergraduate tuition</td>
<td>$10,099.60 ($4,745.21)</td>
<td>$9,124.00</td>
</tr>
<tr>
<td>Undergraduate fees</td>
<td>$860.00 ($299.00)</td>
<td>$683.00</td>
</tr>
</tbody>
</table>

might either not have or be unwilling to provide email addresses for the three external groups (students’ parents, alumni, donors), so institutions were provided the option to send a survey link to these groups themselves. All participating institutions chose to send the links themselves; unfortunately, this meant not having data on how many people received invitations to participate. Institutions were given survey invitation emails to send in late January of 2013. The email included a brief explanation of the project, with a link to the online survey. A follow-up (reminder) email was provided to be sent one week later, in early February 2013. Data collected in Qualtrics were downloaded for analysis.

In all, 361 individuals accessed the survey, although some did not respond to the items related to an accountability template, leaving 274 meaningful responses. Table 16 displays the number and percentage of respondents by missing-data category, and Table 17 presents

Table 16

Phase Two Respondent Count and Percentage by Missing Data Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responded to demographics only</td>
<td>20</td>
<td>5.54%</td>
</tr>
<tr>
<td>Did not respond to any item</td>
<td>67</td>
<td>18.56%</td>
</tr>
<tr>
<td>Responded to at least 1 template-related item</td>
<td>274</td>
<td>75.90%</td>
</tr>
</tbody>
</table>
demographic data for respondents. Figure 5 shows demographic information for respondents. Of the 274 included respondents, 188 (69.0%) came from one institution. Compounded with the small number of institutions participating, this disparate percentage from a single institution calls into question the representativeness of the sample and therefore of the results. As with the number of institutions, so the number of respondents from most participating institutions was disappointing and limiting.

Table 17

*Phase Two Respondent Demographic Percentages*

<table>
<thead>
<tr>
<th>Demographic area</th>
<th>%</th>
<th>(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>46.4</td>
<td>(127)</td>
</tr>
<tr>
<td>Male</td>
<td>53.3</td>
<td>(146)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.7</td>
<td>(2)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.7</td>
<td>(2)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0.7</td>
<td>(2)</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1.8</td>
<td>(5)</td>
</tr>
<tr>
<td>White</td>
<td>93.4</td>
<td>(256)</td>
</tr>
<tr>
<td>Other</td>
<td>1.8</td>
<td>(5)</td>
</tr>
<tr>
<td><strong>Highest degree</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 4-year college</td>
<td>34.3</td>
<td>(94)</td>
</tr>
<tr>
<td>4-year college</td>
<td>35.0</td>
<td>(96)</td>
</tr>
<tr>
<td>Master’s</td>
<td>24.1</td>
<td>(66)</td>
</tr>
<tr>
<td>Doctorate or professional</td>
<td>6.2</td>
<td>(17)</td>
</tr>
<tr>
<td><strong>Stakeholder group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrator</td>
<td>3.6</td>
<td>(10)</td>
</tr>
<tr>
<td>Alumni</td>
<td>31.8</td>
<td>(87)</td>
</tr>
<tr>
<td>Donor</td>
<td>27.0</td>
<td>(74)</td>
</tr>
<tr>
<td>Faculty</td>
<td>9.1</td>
<td>(25)</td>
</tr>
<tr>
<td>Parent</td>
<td>8.4</td>
<td>(23)</td>
</tr>
<tr>
<td>Staff</td>
<td>5.1</td>
<td>(14)</td>
</tr>
<tr>
<td>Student</td>
<td>13.1</td>
<td>(36)</td>
</tr>
</tbody>
</table>
**Data analysis.** Initial data review and screening were completed in PASW 18. Differences among (a) institutions and (b) respondent categories were anticipated; due to the disparity in respondent count across institutions, differences among institutions were not evaluated.

Median importance levels for each metric were calculated. To include all metrics with at least moderate support, any metric having a median importance greater than four (on the seven-point scale) was included in the template.

Differences in importance among stakeholder groups were examined using MANOVA. Differences, if found, would have informed the placement of metrics within the template. This would not have resulted in templates that were adaptive in Spellings Commission chair Charles Miller’s (n.d.) sense of allowing individuals to create personalized weighting systems. But templates could be adaptive by including different screens focused on different stakeholder...
groups (i.e., one for students, one for alumni, etc.), each of which prioritized the information most valued by that group.

**Variables in sample profiles.** Beyond basic template design, this phase included preparation of specific profiles for fictional colleges, to enable Phase Three. As the profiles were created, levels of the accountability data and the institutional demographics were manipulated. Four institutional profiles were created using mock data, one in each index category (At-risk, Challenged, Effective, Excelling). Institutional levels on each template metric were developed to align with its index category.

**Tools and rationale.** Profiles were created using Google Sites and may be viewed online at https://sites.google.com/a/gbs.edu/dissertationsites/. This was done for two reasons. First, Google Sites is a free system, so institutions would have no additional cost burden if they adopted it for creating institutional profiles. Second, it easily interfaces with Google Documents and Google Spreadsheets (including graphs), making it convenient to add and display accountability information.

**Results**

Initial data review and screening was completed in PASW 18. Metrics having a median importance greater than four (on the seven-point scale) were identified for inclusion in the template. As can be seen in Table 18, all presented metrics scored higher than four, so all were included; this supports Hypothesis 6, which stated that stakeholders would support including all presented information types.

Given cultural trends toward more information and data for consumers, stakeholders’ preference to include all presented metrics offers no surprise. As with experts in Phase One, there seems to be a bias toward more information. For experts, this was tempered by lack of
knowledge concerning metrics, but it appears that stakeholders in general may default to including information broadly.

It seems somewhat surprising that stakeholders rated student-experience information (SSI and NSSE scores) as less important (median=5 out of 7) than other metrics; the lower importance for anonymous samples of student work (also 5) seems less surprising, since such information would require stakeholder “processing” to be of much use in evaluating the institutions. It may be that student experience is of less value in the context of the specific question, which was what information respondents need in order to evaluate an institution’s fulfillment of its mission.

Table 18

Means, Standard Deviations, Medians and Interquartile Ranges for Importance Scores for Metrics on a 7-point Scale

<table>
<thead>
<tr>
<th>Metric</th>
<th>M (SD)</th>
<th>Median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuition/fees</td>
<td>6.31 (0.746)</td>
<td>6</td>
<td>6-7</td>
</tr>
<tr>
<td>Room/board costs</td>
<td>6.10 (0.843)</td>
<td>6</td>
<td>6-7</td>
</tr>
<tr>
<td>Student testimonials/biographies</td>
<td>6.01 (0.922)</td>
<td>6</td>
<td>5-7</td>
</tr>
<tr>
<td>Alumni testimonials/biographies</td>
<td>6.00 (0.870)</td>
<td>6</td>
<td>5-7</td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
<td>6.00 (0.866)</td>
<td>6</td>
<td>5-7</td>
</tr>
<tr>
<td>Completion rate</td>
<td>5.94 (0.871)</td>
<td>6</td>
<td>5-7</td>
</tr>
<tr>
<td>Retention rate</td>
<td>5.91 (0.867)</td>
<td>6</td>
<td>5-7</td>
</tr>
<tr>
<td>General Education test</td>
<td>5.85 (0.888)</td>
<td>6</td>
<td>5-6</td>
</tr>
<tr>
<td>Faculty evaluations of student work</td>
<td>5.64 (0.928)</td>
<td>6</td>
<td>5-6</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td>5.66 (0.920)</td>
<td>6</td>
<td>5-6</td>
</tr>
<tr>
<td>SSI scale scores</td>
<td>5.51 (0.857)</td>
<td>5</td>
<td>5-6</td>
</tr>
<tr>
<td>NSSE scale scores</td>
<td>5.22 (0.967)</td>
<td>5</td>
<td>5-6</td>
</tr>
<tr>
<td>Anonymous samples of student work</td>
<td>4.81 (1.326)</td>
<td>5</td>
<td>4-6</td>
</tr>
</tbody>
</table>
Differences in importance among stakeholder groups were examined using a one-way MANOVA design with thirteen dependent variables, the thirteen metrics listed above.\(^\text{18}\) The independent variable was stakeholder group (administrator, alumnus/alumna, donor, faculty member, parent of a current student, staff member, student).

**Outliers.** Univariate outliers were identified by examination of \(z\) scores. A total of 31 respondents were found to have extreme low scores (less than -3.29, \(p<.001\)) on dependent variables; these respondents were removed from the data set prior to running the MANOVA.\(^\text{19}\) Table 19 shows comparative demographics for the removed outliers and the retained responses. The differences in gender balance and ethnicity representation are cause for some concern, as female and non-White respondents may tend toward a different perspective that was for analytical purposes seen as outlying; this possible demographic difference in satisfaction presents a possibility for future research. Screening for multivariate outliers was done before and after removal of univariate outliers; after univariate outliers were removed, only six respondents were considered multivariate outliers based on Mahalanobis’s distance (distance greater than 36.123, with \(df=14\)). As there were few multivariate outliers, and only two of those had distances greater than 40, they were not seen as sufficiently removed from the rest of the distribution as to represent serious threats to the MANOVA’s validity; in order to retain as much data as possible and to represent legitimate minority views, all six were retained.

**Normality.** Of the thirteen dependent variables, five were found to have unacceptably large (greater than \(\pm3.29\)) deviations from normality when standardized values (\(z\) scores) for

---

\(^\text{18}\) Recall that experts had eliminated some metrics, seeing them as not important to include. Additionally, experts were not asked about student or alumni testimonials, since expert opinion was used to develop a quantitative index, into which narrative would not fit. This explains the different number of metrics discussed, 13 for stakeholders and 9 for experts.

\(^\text{19}\) Screening for univariate outliers was conducted twice, before and after transformations were performed (see under Normality). If univariate outliers were left during transformation, the number of outliers needing to be removed increased from 31 to 37. For this reason, univariate outliers were removed prior to transformation.
Skewness and kurtosis were examined. Table 20 reports these five variables along with skewness and kurtosis data, both before and after square-root transformation; transformations were completed for all groups, but only groups with excessive deviations from normality are included in the table. For all variables and all groups, square-root transformations resulted in skewness and kurtosis $z$ scores below $\pm 3.29$.

Table 19

Demographic Percentages ($n$) for Removed Outliers and Retained Respondents

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Removed</th>
<th>Retained</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Institution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.85% (1)</td>
<td>0.87% (2)</td>
</tr>
<tr>
<td>2</td>
<td>76.92% (20)</td>
<td>79.48% (182)</td>
</tr>
<tr>
<td>3</td>
<td>15.38% (4)</td>
<td>10.92% (25)</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>6.11% (14)</td>
</tr>
<tr>
<td>6</td>
<td>–</td>
<td>0.87% (2)</td>
</tr>
<tr>
<td>7</td>
<td>3.85% (1)</td>
<td>1.75% (4)</td>
</tr>
<tr>
<td><strong>Role</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrator</td>
<td>–</td>
<td>4.26% (11)</td>
</tr>
<tr>
<td>Alumnus/a</td>
<td>36.67% (11)</td>
<td>33.72% (87)</td>
</tr>
<tr>
<td>Donor</td>
<td>30% (9)</td>
<td>25.97% (67)</td>
</tr>
<tr>
<td>Faculty</td>
<td>6.67% (2)</td>
<td>9.3% (24)</td>
</tr>
<tr>
<td>Parent</td>
<td>13.33% (4)</td>
<td>7.36% (19)</td>
</tr>
<tr>
<td>Staff</td>
<td>6.67% (2)</td>
<td>4.65% (12)</td>
</tr>
<tr>
<td>Student</td>
<td>6.67% (2)</td>
<td>14.73% (38)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>53.13% (17)</td>
<td>46.33% (120)</td>
</tr>
<tr>
<td>Male</td>
<td>46.88% (15)</td>
<td>53.67% (139)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian/Alaskan Native</td>
<td>3.13% (1)</td>
<td>0.39% (1)</td>
</tr>
<tr>
<td>Asian</td>
<td>3.13% (1)</td>
<td>0.39% (1)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>–</td>
<td>0.77% (2)</td>
</tr>
<tr>
<td>Latino</td>
<td>3.13% (1)</td>
<td>1.54% (4)</td>
</tr>
<tr>
<td>White</td>
<td>87.5% (28)</td>
<td>95.37% (247)</td>
</tr>
<tr>
<td>Other</td>
<td>3.13% (1)</td>
<td>1.54% (4)</td>
</tr>
</tbody>
</table>
Table 20

Skewness and Kurtosis before and after Square-root Transformation

<table>
<thead>
<tr>
<th>Metric (Group)</th>
<th>Skew Before (z)</th>
<th>Skew After (z)</th>
<th>Kurtosis Before (z)</th>
<th>Kurtosis After (z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bible Content Test (Alumni)</td>
<td>-1.14 (-4.06)</td>
<td>0.68 (2.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completion rate (Alumni)</td>
<td></td>
<td></td>
<td>2.16 (3.82)</td>
<td>0.50 (0.88)</td>
</tr>
<tr>
<td>Tuition/fees (Alumni)</td>
<td>0.96 (-3.43)</td>
<td>0.55 (1.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSSE scale scores (Donors)</td>
<td>-1.24 (-3.96)</td>
<td>0.80 (2.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student work samples (Alumni)</td>
<td>-0.96 (-3.36)</td>
<td>0.29 (1.03)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Linearity and multicollinearity.** Examination of bivariate scatterplots within each group show that deviations from normality, although within the tolerances described above, hinder linearity among dependent variable pairs, although curvilinearity does not appear to be present. Transformations can be used to improve linearity in dependent variables (Tabachnick & Fidell, 2007), but these have already been done, as discussed above.

Multicollinearity was checked by Box’s $M: M=574.286, p=.005$. Given the test’s sensitivity, a conservative interpretation is recommended; as the result was not significant at the $p<.001$ level, MANOVA proceeded, but Pillai’s trace was used as it is more appropriate given imperfect alignment with assumptions (Tabachnick & Fidell, 2007).

**MANOVA.** Using Pillai’s trace, the combined dependent variables (metric importance ratings) were not significantly different across stakeholder roles, $F(78, 1140)=1.162, p=0.165$, with observed power of 1.000 (see Table 21). This result supports Hypothesis 7, which stated that there would not be significant differences among groups’ importance ratings. The lack of significant difference in importance ratings across stakeholder groups lends support for programs, such as the VSA, that use standardized templates for public accountability. (There was one significant difference across education levels, but all groups still had median ratings high enough to include the metric.) Because of this result, online templates were developed around
Table 21

*Multivariate Analysis of Variance of Combined Dependent Variables by Respondent Role*

<table>
<thead>
<tr>
<th>Source</th>
<th>Pillai’s Trace</th>
<th>Multivariate F</th>
<th>df₁</th>
<th>df₂</th>
<th>p</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role</td>
<td>0.451</td>
<td>1.187</td>
<td>78</td>
<td>1140</td>
<td>0.133</td>
<td>1.000</td>
</tr>
</tbody>
</table>

information categories, rather than around respondent group.

**Phase Two Summary**

Surprisingly, Phase Two found no significant differences among stakeholder groups’ perceived information needs. This led to development of an accountability template organized not around stakeholder group (e.g., “information for students,” “information for alumni”) but around information type (costs, student learning/academics, student success, student experience). This template was used to create four mock profiles, which enabled Phase Three’s research into stakeholder understanding of and satisfaction with the accountability template.
Chapter Five: Phase Three: Stakeholder Evaluation of Institutional Profiles

Repeated searches of the literature and of transparency-initiative websites revealed no empirical studies of stakeholder response to/satisfaction with systems intended to hold institutions accountable in a transparent fashion. But of course if stakeholders cannot understand these systems, or find their information insufficient or otherwise unsatisfactory, then the systems fail to accomplish a key goal, satisfying stakeholder information needs. Phase Three addressed exactly this issue. The same groups from Phase Two – e.g., students, parents, alumni, faculty – were included in Phase Three, but this does not necessarily mean that the same individuals responded; since institutions distributed the links, institutions may have sent them to different individuals, or different individuals may have chosen to respond to Phase Three than responded to Phase Two. Stakeholder groups were asked to view online accountability profiles for four imaginary Bible colleges. Then, stakeholders answered factual questions about the profiles to measure stakeholder accuracy of interpretation, followed by items to measure stakeholder satisfaction with the metrics and online profiles.

Methods

Research design and rationale. To measure stakeholder satisfaction with a proposed online transparency system within the ABHE context, respondents, who were stakeholders at ABHE institutions, were asked to complete a survey. The survey measured (a) respondent understanding of profile information, (b) respondent satisfaction with profile content and (c) respondent satisfaction with profile design. Respondents reviewed four (fictional) institutional profiles as they completed the survey.

A survey was used as the most appropriate means to collect large-scale stakeholder response, although the scale was not as large as was hoped. While some issues – especially
stakeholder understanding and design-related information – could more richly be measured with focus groups or interviews, this study’s focus was on broad usefulness of online profiles. The smaller, more qualitative approaches were seen as not enabling conclusions on these matters. Narrower studies could be implemented following this one, more focused on specific user interaction with profiles and finer-grained understanding of information.

**Population.** The population here comprised the same stakeholder groups as in Phase Two: administrators, staff, faculty, students, current students’ parents, alumni, donors. (It is worth noting that this also included most Phase One participants, who are employees of member institutions.) The target population, then, included all individuals in these categories at all ABHE-affiliated institutions.

**Survey design.** The survey was designed and administered using Qualtrics, an online survey system. As in Phase Two, the online survey allowed for the broadest possible reach; it was anticipated that the time required to complete the survey would be 30 minutes or less. Following an informed-consent page, the survey had four basic sections. (The survey itself included a link to the website containing the four institutional profiles, which opened in a new window for respondent convenience.) The survey is provided in Appendix D.

**Respondent demographics.** To enable analysis, respondents were asked to indicate (a) their role (administrator, staff, faculty, student, student’s parent, alumnus/alumna, donor) and (b) their institutional affiliation (from a list of institutions participating in the study). Gender, age, ethnicity and education level were also collected.

**Respondent understanding.** First, respondents answered a series of questions measuring their ability to understand the information presented in the profiles. To reduce “survey fatigue” by limiting the number of questions presented to a respondent and thereby shortening the survey
(Galesic & Bosnjak, 2009; Moss & Hendry, 2002; Swanson & Holton, 2005), only some of the questions within a cluster were presented randomly (see Appendix D). Clusters covered the different kinds of metrics included. For example, respondents were randomly shown items in two of the following four learning-related areas: General Education student learning; Bible student learning; faculty evaluation of student work; faculty degrees. Randomization also occurred within the other three areas: retention/completion, costs and student experience.

For each item, respondents were asked to rate their confidence in their answer as a proxy for effort; this information was used to weight respondents’ correct answers. From these items, a mean weighted understanding score was developed for each respondent.

Along with the understanding questions, these pages included a timer item. Qualtrics tracked the amount of time respondents spend on a page, as well as the number of “clicks” on the page. Time-on-page was checked during data analysis as an indicator of respondent effort to answer items correctly.

**Respondent satisfaction.** Respondents were asked about their satisfaction with specific information presented as well as with information categories. Information satisfaction was measured using a 7-point Likert-type item for each aspect of the transparency template, ranging from 1=“Not at all satisfied” to 7=“Very satisfied.” A mean information-satisfaction score was calculated for each respondent.

**Sampling and participant recruitment.** Participant recruitment for Phase Three took place during Phase Two: the same institutions were used for Phase Three. Phase-three survey invitations were provided to institutions: the email included a brief explanation of the project, with a link to the online survey. A follow-up (reminder) email was provided to be sent one week later. Perhaps because the phase-three survey followed Phase Two by roughly one week (mid-
February 2013), there were fewer respondents. Of the 144 respondents, 67 provided substantive responses (answering at least two objective items and at least two satisfaction items). Table 22 shows respondent counts and percentages for various kinds of missing data. In order to maximize use of the data, no responses were removed from analysis due to missing values; this means that different analyses had different counts.

In this case, five institutions were represented, but again a majority of responses (75.4%) were from a single institution, the same institution heavily represented in Phase Two. Table 23 displays demographics for the 67 substantive respondents.

**Data collection.** The survey collected mostly numeric information using only closed items. Data collected in Qualtrics was downloaded for analysis.

**Data analysis.** Initial data review and screening were completed in PASW 18. Differences among (a) institutions, and (b) respondent categories were anticipated. Plans called for data to be analyzed hierarchically using HLM software and employing the two-level hierarchical linear modeling (HLM2) approach (Garson, 2012; Tabachnick & Fidell, 2007) in two separate models, one with accuracy as the dependent (outcome) variable and one with satisfaction as the dependent (outcome) variable. Independent (predictor) variables in both models would include individual characteristics (stakeholder group, age, gender, ethnicity, education level) and institutional characteristics (size, region, highest degree level granted, ABHE accreditation status); the satisfaction model added the individual respondent’s accuracy score as another independent (predictor) variable. The model structure is depicted in Figure 6; both models were to use the same structure. In Figure 6, the first subscript indicates association with a particular college and the second subscript identifies an individual respondent; for
Table 22

*Phase Three Missing Data Categories with Respondent Count and Percentage*

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>No responses at all</td>
<td>14</td>
<td>9.72%</td>
</tr>
<tr>
<td>Only demographics</td>
<td>53</td>
<td>36.81%</td>
</tr>
<tr>
<td>One or fewer items</td>
<td>3</td>
<td>2.08%</td>
</tr>
<tr>
<td>Objective only</td>
<td>7</td>
<td>4.86%</td>
</tr>
<tr>
<td>Responses to 2 or more in each section</td>
<td>67</td>
<td>46.53%</td>
</tr>
</tbody>
</table>

Table 23

*Phase Three Respondent Demographic Percentages*

<table>
<thead>
<tr>
<th>Demographic area</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>31</td>
<td>46.27%</td>
</tr>
<tr>
<td>Male</td>
<td>35</td>
<td>52.24%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1</td>
<td>1.49%</td>
</tr>
<tr>
<td>White</td>
<td>63</td>
<td>94.03%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1.49%</td>
</tr>
<tr>
<td>Highest degree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 4-year college</td>
<td>21</td>
<td>31.34%</td>
</tr>
<tr>
<td>4-year college</td>
<td>13</td>
<td>19.40%</td>
</tr>
<tr>
<td>Master’s</td>
<td>24</td>
<td>35.82%</td>
</tr>
<tr>
<td>Doctorate or professional</td>
<td>8</td>
<td>11.94%</td>
</tr>
<tr>
<td>Stakeholder group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrator</td>
<td>5</td>
<td>7.46%</td>
</tr>
<tr>
<td>Alumni</td>
<td>18</td>
<td>26.87%</td>
</tr>
<tr>
<td>Donor</td>
<td>18</td>
<td>26.87%</td>
</tr>
<tr>
<td>Faculty</td>
<td>8</td>
<td>11.94%</td>
</tr>
<tr>
<td>Parent</td>
<td>7</td>
<td>10.45%</td>
</tr>
<tr>
<td>Staff</td>
<td>3</td>
<td>4.48%</td>
</tr>
<tr>
<td>Student</td>
<td>7</td>
<td>10.45%</td>
</tr>
</tbody>
</table>

Example, Student17 would be a student at College 1 and would be the seventh student respondent in that college’s student group. Due to lack of data and small ICC values, ANOVA and ANCOVA were used instead.
Results

Data preparation. Prior to analysis of stakeholder accuracy in template interpretation and of stakeholder satisfaction with the templates, stakeholder responses to factual questions were scored. Table 24 provides the percentage of respondents correctly answering factual questions by metric, across stakeholder groups. Correct-response information, by group, is presented in Table 25 and visualized in Figure 7. In Figure 7, the horizontal axis presents the metrics used in the profiles, and the vertical axis presents the percentage of respondents in each stakeholder group who correctly answered items about the corresponding metric. The size of each “bubble” represents the respondent count for that stakeholder group on items related to that...
metric. This allows comparisons among stakeholder groups across metrics as well as within a metric: for example, the donor group tends to have low accuracy across metrics, while the student group’s accuracy varies substantially from metric to metric.

It may be of some value to consider apparent overall interpretive difficulty with some types of data (see Table 24 and Figure 8). In Figure 8, bubble size represents the number of respondents who answered items about the specified metric. The vertical axis represents the percentage of respondents answering metric-specific items correctly; for ease of viewing, it ranges from 35%-100%. (The horizontal axis is not a meaningful scale, but simply spreads the bubbles for better visibility.) Figure 8 allows for better comparison of accuracies from metric to metric, ignoring stakeholder groups. Across all stakeholder categories, only tuition and fee

Table 24

Percentage of Correct Responses by Metric (All Stakeholder Categories)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Average % correct (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuition/fees</td>
<td>100.00 (20)</td>
</tr>
<tr>
<td>SSI scale scores</td>
<td>76.32 (38)</td>
</tr>
<tr>
<td>General Education test</td>
<td>75.00 (36)</td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
<td>71.43 (35)</td>
</tr>
<tr>
<td>Retention rate</td>
<td>68.42 (38)</td>
</tr>
<tr>
<td>Room/board costs</td>
<td>63.16 (38)</td>
</tr>
<tr>
<td>Completion rate</td>
<td>48.57 (35)</td>
</tr>
<tr>
<td>Faculty evaluations of student work</td>
<td>43.24 (37)</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td>41.18 (34)</td>
</tr>
<tr>
<td>NSSE scale scores</td>
<td>39.39 (33)</td>
</tr>
</tbody>
</table>
Figure 7. Accuracy by stakeholder group, with bubble size corresponding to respondent count

Figure 8. Percentage correct across groups, with bubbles sized by respondent count
information was correctly understood by 100% of respondents (n=20); in contrast, room/board costs were understood by 63.16% (n=38). Most standardized instruments (Bible test, n=35; General Education test, n=36; the SSI, n=38) show reasonable levels of correct understanding, with average percentages in the 70s; in contrast, NSSE scale scores, which from Phase One seems perhaps lesser known by those involved in Biblical higher education, were correctly interpreted by only 39.39% of respondents (n=33). Faculty evaluations of student work – something not favored by expert evaluators – was understood by less than half of respondents (43.24%, n=37), and percentage of faculty with doctoral degrees, a common measure, was also understood by less than half (41.18%, n=34). Perhaps most challenging results for thinking about accountability is the mediocre accuracy with understanding completion (48.57%, n=35) and retention (68.42%, n=38) rates. Opinion from most perspectives on higher-education accountability favors use of these numbers. As Table 25 shows, scores were low across many respondent groups for these items.

Various possible explanations for low rates of accurate understanding were considered. While time-on-page from the survey shows a few significant correlations with correct answers, these are too few to account for differences (see significant correlations marked in Table 25). The type of cognitive process required for each survey item was also considered: some items required comparing across the four institutional profiles, while some required comparing results within a single table on each profile. Still, no pattern emerged, as similar items showed both high and low accuracy (e.g., tuition/fees and percentage of faculty with doctoral degrees were very similar items; the first had 100% accuracy, the second 41.18% accuracy).

For each accuracy item, respondents were asked to rate their confidence in their responses; confidence values were used as weights to calculate weighted accuracy scores.
Table 25

*Percentage of Respondents Correctly Answering Template-understanding Questions by Metric and by Stakeholder Category*

<table>
<thead>
<tr>
<th>Metric</th>
<th>Administrators</th>
<th>Alumni</th>
<th>Donors</th>
<th>Faculty</th>
<th>Parents</th>
<th>Staff</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>General Education test</td>
<td>66.7</td>
<td>3</td>
<td>80</td>
<td>10</td>
<td>50</td>
<td>6</td>
<td>66.7</td>
</tr>
<tr>
<td>ABHE Bible Content Test</td>
<td>100</td>
<td>4</td>
<td>75</td>
<td>8</td>
<td>37.5</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>Faculty evaluations of student work</td>
<td>50**</td>
<td>2</td>
<td>44.4</td>
<td>9</td>
<td>18.2</td>
<td>11</td>
<td>80</td>
</tr>
<tr>
<td>Percentage of faculty with doctoral degrees</td>
<td>50**</td>
<td>2</td>
<td>45.5*</td>
<td>11</td>
<td>33.3*</td>
<td>9</td>
<td>33.3</td>
</tr>
<tr>
<td>Completion rate</td>
<td>100</td>
<td>2</td>
<td>50</td>
<td>12</td>
<td>25</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Retention rate</td>
<td>100</td>
<td>3</td>
<td>87.5</td>
<td>8</td>
<td>36.4</td>
<td>11</td>
<td>33.3</td>
</tr>
<tr>
<td>Tuition/fees</td>
<td>100</td>
<td>2</td>
<td>100</td>
<td>5</td>
<td>100</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>Room/board costs</td>
<td>100</td>
<td>3</td>
<td>72.7</td>
<td>11</td>
<td>27.3</td>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>SSI scale scores</td>
<td>100</td>
<td>5</td>
<td>57.1</td>
<td>7</td>
<td>50</td>
<td>8</td>
<td>83.3</td>
</tr>
<tr>
<td>NSSE scale scores</td>
<td>--</td>
<td>0</td>
<td>46.2</td>
<td>13</td>
<td>20</td>
<td>10</td>
<td>66.7</td>
</tr>
<tr>
<td>Average weighted %</td>
<td>88.47</td>
<td>62.77</td>
<td>33.34</td>
<td>63.63</td>
<td>85.75</td>
<td>86.67</td>
<td>69.04</td>
</tr>
</tbody>
</table>

*Note: Asterisks show results where accuracy correlated (point-biserial correlation) with time-on-page as recorded by the survey system. *p<0.05; **p<0.01.*
These scores were then averaged to create an average weighted accuracy for each respondent. Table 26 provides descriptive statistics for this average weighted accuracy variable. Confidence weighting does not substantially alter the relative performance of the stakeholder groups: average weighted accuracy (from Table 26) correlates strongly with weighted percentage correct (from Table 25), $r=0.97$, $p=0.0002$.

Average satisfaction scores were calculated by averaging respondents’ reported satisfaction levels for each metric. Resulting average satisfaction scores for each stakeholder group are reported in Table 27. Interestingly, average satisfaction for stakeholder groups does not correlate strongly with average weighted accuracy, $r=0.51$, $p=0.245$. In other words, there was not a clear or strong relationship between understanding of the profile information and satisfaction with the metrics.

**Hierarchical linear modeling.** Following data preparation, HLM software was used to run the unconditional model for the two-level hierarchical design presented in Figure 6, above. Response counts were minimally sufficient, but solutions were derived for two unconditional models, one with average accuracy as the outcome and a second with average satisfaction as the outcome. Results from the unconditional models were used to calculate intraclass correlation coefficients (ICC$s$) for the accuracy model (ICC=0.012) and for the global satisfaction model (ICC=0.0001). With such small ICC$s$, HLM analysis was unnecessary, so ANOVA/ANCOVA designs were used instead.

**Accuracy ANOVA.** A one-way between-subjects analysis of variance was performed to examine differences in average weighted accuracy across the seven stakeholder groups. The

---

20 Here and in other places parametric methods are used with Likert-type data. Cooley (2011) has extensively reviewed the decades-long discussion in the literature about parametric procedures with Likert-type (ordinal) data. Garson (2013) notes that “in practice Likert items… are very commonly used with interval procedures in social science, provided the scale item has at least 5 and preferably 7 categories” (p. 62). Norman (2010) argues from theory and from studies of robustness that Likert data can be used with parametric approaches.
Table 26

Stakeholder Group Means for Average Weighted Accuracy

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrator</td>
<td>6</td>
<td>3.63</td>
<td>1.75</td>
<td>[1.62, 5.30]</td>
</tr>
<tr>
<td>Alumnus/Alumna</td>
<td>20</td>
<td>2.52</td>
<td>1.80</td>
<td>[1.67, 3.36]</td>
</tr>
<tr>
<td>Donor</td>
<td>19</td>
<td>1.34</td>
<td>1.47</td>
<td>[0.63, 2.05]</td>
</tr>
<tr>
<td>Faculty member</td>
<td>10</td>
<td>2.54</td>
<td>1.74</td>
<td>[1.29, 3.78]</td>
</tr>
<tr>
<td>Parent of a current student</td>
<td>7</td>
<td>4.14</td>
<td>1.21</td>
<td>[3.02, 5.27]</td>
</tr>
<tr>
<td>Staff member</td>
<td>3</td>
<td>4.13</td>
<td>1.15</td>
<td>[1.26, 7.00]</td>
</tr>
<tr>
<td>Student</td>
<td>9</td>
<td>2.84</td>
<td>1.84</td>
<td>[1.42, 4.25]</td>
</tr>
</tbody>
</table>

Note: Means that do not share any subscripts differ significantly from each other by Tukey’s HSD at p<0.05. Thus, the Donor group mean differs significantly from the group means for Parents and Staff members.

Table 27

Stakeholder Group Means for Average Satisfaction with Metrics

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrator</td>
<td>5</td>
<td>4.85</td>
<td>0.45</td>
<td>[4.29, 5.40]</td>
</tr>
<tr>
<td>Alumnus/Alumna</td>
<td>18</td>
<td>4.86</td>
<td>0.88</td>
<td>[4.42, 5.29]</td>
</tr>
<tr>
<td>Donor</td>
<td>18</td>
<td>4.75</td>
<td>1.33</td>
<td>[4.09, 5.41]</td>
</tr>
<tr>
<td>Faculty member</td>
<td>8</td>
<td>4.78</td>
<td>0.95</td>
<td>[3.98, 5.58]</td>
</tr>
<tr>
<td>Parent of a current student</td>
<td>7</td>
<td>4.77</td>
<td>0.99</td>
<td>[3.86, 5.68]</td>
</tr>
<tr>
<td>Staff member</td>
<td>3</td>
<td>5.59</td>
<td>1.27</td>
<td>[2.43, 8.76]</td>
</tr>
<tr>
<td>Student</td>
<td>7</td>
<td>4.91</td>
<td>1.30</td>
<td>[3.70, 6.12]</td>
</tr>
</tbody>
</table>

assumption of homogeneity of variance was met (Tabachnick & Fidell, 2007), and no outliers were identified. While skewness and kurtosis statistics for all groups had z-scores lower than ±3.1 (Tabachnick & Fidell, 2007, pp. 79–80), histograms revealed some deviation from normality, which can increase the risk of a Type I error (Lantz, 2013).

Null Hypothesis 8 states that there are not significant differences among stakeholder groups’ (average weighted) accuracy in interpreting data presented in online institutional transparency profiles. The ANOVA tests this. Average weighted accuracy differences among
stakeholder groups were statistically significant, with $F(6, 67)=3.311, p=.006$; see Table 28. Noting that the nonnormality can increase the Type I error rate, the $p$-value obtained here was sufficiently small to reject Null Hypothesis 8 at the $\alpha=0.05$ level.

**Significant differences between groups.** SPSS was used to produce Tukey’s HSD post-hoc tests for between-group differences in average weighted accuracy. The donor group was found to have significantly lower accuracy than two other groups, the parent group and the staff group, at the $p<0.05$ level; see Table 26.

**Satisfaction ANCOVA.** A one-way between-subjects analysis of covariance was performed to examine differences in average satisfaction with metrics across the seven stakeholder groups. The average weighted accuracy score was the covariate. The assumption of linearity appears not to be met for grouped data, although examination of scatterplots within groups appears more acceptable; this lack of linearity results in a more conservative test (Tabachnick & Fidell, 2007, p. 202). Average satisfaction showed levels of skewness and kurtosis greater than $z=|3.1|$ for grouped data (and for staff skewness), but transformations exacerbated these values. Except for staff skewness ($n=3$), skewness and kurtosis levels within groups met expectations (Tabachnick & Fidell, 2007, pp. 79–80). Tests of the assumptions of homogeneity of variance and homogeneity of regression found no violation of the assumptions (Tabachnick & Fidell, 2007), and no outliers were identified. There are sufficient challenges to test assumptions to warrant care in interpreting test results.

After adjustment by the covariate, average satisfaction did not vary significantly by stakeholder group, $F(6, 58)=0.226, p>0.05$; see Table 29.
Hypothesis 9 stated that, on average, all stakeholder groups would be satisfied that institutional transparency profiles provide sufficient information to enable viewers to evaluate institutional health, with satisfaction defined as a mean satisfaction score higher than neutral (4). As Table 27 shows, Hypothesis 9 was supported by the results, with all groups having average satisfaction greater than 4. It should be noted, though, that four groups – faculty, parents, staff, students – had sufficient variance that the 95% confidence interval included values below 4, raising questions about this conclusion.

The only slightly positive results suggest that stakeholders may not know what they want: Phase Two’s results showed high support for these very same metrics (see Table 18), and it seems reasonable to think that stakeholder satisfaction levels should be similar to stakeholder
importance levels. Yet independent-measures $t$-tests\textsuperscript{21} for aggregate data from phases two and three show significant differences, $p<0.001$, for all metrics except anonymous samples of student work. Again, while the sample here is small, results suggest that stakeholders like the idea of accountability data, but seeing the data proves less satisfactory to them. Given the generally low accuracy levels, it may be that respondents’ difficulty in interpreting the data contributes to reduced satisfaction, although none of the tests performed found a consistent relationship between accuracy and satisfaction.\textsuperscript{22}

**Template satisfaction ANOVA.** Hypothesis 10 stated that there would be significant differences among stakeholder groups’ level of satisfaction with online institutional transparency templates, measured by average global satisfaction. Global satisfaction was measured by averaging two items, one of which measured respondents’ satisfaction with student-learning information and one of which measured respondents’ satisfaction with institutional-cost information. Descriptive statistics by stakeholder group are shown in Table 25. Interestingly, stakeholder groups’ global satisfaction (Table 30) did not neatly correlate with the same groups’ average satisfaction across metrics (Table 27), $r=0.248$, $p=0.59$. Three groups (alumni, donors, faculty) had slightly higher ($M=0.04$) global satisfaction, while the other four groups (administrators, parents, staff, students) had lower global satisfaction ($M=-0.30$). Were there a clear pattern, it would suggest that a transparency system had higher or lower satisfaction than the individual measures, but that was not consistently the case.

Global satisfaction showed a negative correlation with average weighted accuracy ($r=-0.46$, $p=0.30$) and with average percent correct ($r=-0.35$, $p=0.44$). As with metric satisfaction,\textsuperscript{21} $F$-tests showed unequal variance for all except the anonymous samples of student work; for all other metrics, Welch’s $t$-test was used.\textsuperscript{22} Significant correlations ($p<0.05$) were found between accuracy and satisfaction for SSI and NSSE items, but the correlation for NSSE was negative, and both SSI and NSSE accuracy had stronger correlations with satisfaction on unrelated items, calling into question the correlations’ meaningfulness.

\textsuperscript{21} $F$-tests showed unequal variance for all except the anonymous samples of student work; for all other metrics, Welch’s $t$-test was used.

\textsuperscript{22} Significant correlations ($p<0.05$) were found between accuracy and satisfaction for SSI and NSSE items, but the correlation for NSSE was negative, and both SSI and NSSE accuracy had stronger correlations with satisfaction on unrelated items, calling into question the correlations’ meaningfulness.
then, there was not a clear relationship between understanding of the information and satisfaction with the overall transparency template. The problem this poses is that transparent accountability systems may satisfy some stakeholders without actually empowering them to use the information effectively to hold institutions accountable.

A one-way between-subjects analysis of variance was performed to examine differences in average global satisfaction across the seven stakeholder groups. Evaluation of the assumptions of normality and homogeneity of variance were satisfactory, and no outliers were identified.

Table 30

*Stakeholder Group Means on Average Global Satisfaction*

<table>
<thead>
<tr>
<th>Role</th>
<th>n</th>
<th>M (SD)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrator</td>
<td>5</td>
<td>4.80 (0.76)</td>
<td>[3.86, 5.74]</td>
</tr>
<tr>
<td>Alumnus/Alumna</td>
<td>18</td>
<td>4.89 (1.27)</td>
<td>[4.26, 5.52]</td>
</tr>
<tr>
<td>Donor</td>
<td>17</td>
<td>4.79 (1.49)</td>
<td>[4.03, 5.56]</td>
</tr>
<tr>
<td>Faculty member</td>
<td>6</td>
<td>4.83 (1.54)</td>
<td>[3.22, 6.45]</td>
</tr>
<tr>
<td>Parent of a current student</td>
<td>7</td>
<td>4.43 (1.30)</td>
<td>[3.22, 5.64]</td>
</tr>
<tr>
<td>Staff member</td>
<td>3</td>
<td>4.83 (1.53)</td>
<td>[1.04, 8.63]</td>
</tr>
<tr>
<td>Student</td>
<td>7</td>
<td>4.86 (1.44)</td>
<td>[3.53, 6.18]</td>
</tr>
</tbody>
</table>

Table 31

*Results of the Average Global Satisfaction ANOVA*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>1.139</td>
<td>6</td>
<td>0.102</td>
<td>0.996</td>
</tr>
<tr>
<td>Within groups</td>
<td>104.179</td>
<td>56</td>
<td>(1.860)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>105.317</td>
<td>62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: MSE in parentheses.*
Global satisfaction did not vary significantly among stakeholder groups, with $F(6, 56)=0.102, p=0.996$; see Table 31. Thus, Null Hypothesis 10 was not rejected. This result supported the others in this phase, as it highlighted the lack of a clear relationship between understanding and satisfaction. While there were significant differences in understanding among groups, this did not mean satisfaction also varied significantly. Indeed, the two groups with highest global satisfaction (alumni and students) were in the lower part of the accuracy averages (sixth and fourth, out of seven, respectively).

**Phase Three Summary**

Phase Three explored stakeholder understanding of and satisfaction with an online accountability template or system. While Phase Two’s heavy inclusion of stakeholders in template development was a new project, Phase Three’s exploration of stakeholder understanding and satisfaction represented the most important new information in this study. Critically, stakeholder accuracy was not consistently high for any stakeholder group, and only the tuition information was understood well across groups. Additionally, despite Phase Two’s conclusion that stakeholder groups valued the information included in the templates, when they actually reviewed it in the online profiles, stakeholders expressed limited satisfaction with the information on the Phase Three survey.
Chapter Six: Discussion

In this final chapter, results from across all three phases are discussed. Here the intention is to summarize major findings and consider their implications. This includes a review of study limitations, as well as the identification of both research and action opportunities.

Summary of Major Findings

The study’s four objectives provide organization for this summary of key findings. In the summary, study results are related to the literature reviewed in Chapter Two.

Objective One: Multivariate index. Phase One, which included both Objectives One and Two, addressed development of a multivariate index of institutional health (defined in alignment with accreditation as an institution’s ability to fulfill its mission). For Objective One, ABHE team evaluators were surveyed to identify metrics for inclusion in the index, as well as weights for each included metric and cutpoints to categorize an institution as At-risk, Challenged, Effective or Excelling.

Contrary to Hypothesis 1, based on majority support responding ABHE team evaluators did eliminate the following presented metrics from inclusion in a multivariate index of institutional health: faculty evaluation of senior-level students’ work and National Survey of Student Engagement (NSSE) results. These particular metrics, though, had the highest reported levels of “insufficient knowledge to evaluate.” It should be noted that respondents may not have understood what was meant by “Aggregated faculty evaluations of senior-level work based on ABHE rubrics,” as that item was phrased on the survey. But failure to include this item suggests a possible misalignment between ABHE team evaluators and voices in the accountability literature that highly value rubric-driven and portfolio-type assessment (Banta, 2007; Basken, 2008; Eaton, 2008; Peter D. Hart Research Associates, 2008; Schneider, 2009). This may also
reflect the tension identified in the literature between the improvement and accountability paradigms (Ewell, 2008), with team evaluators taking more of an accountability perspective.

The responding evaluators included the following metrics in the index, grouped by principal component: retention and completion rates (component 1); room/board costs, total weighted financial factor, tuition/fee costs (component 2); and percentage of faculty with doctoral degrees, Student Satisfaction Inventory (SSI) scores, ABHE Bible Content Test scores, General Education test scores (component 3).

Based on weights assigned to each metric, Null Hypothesis 2, which stated that evaluators would assign equal weights to all measures, was tentatively rejected, although the small sample (n=31) and missing data limit confidence in this conclusion. Respondents weighted the metrics as follows: retention rate (0.142); room/board costs (0.140); completion rate (0.135); percentage of faculty with doctoral degrees (0.101); total weighted financial factor (0.093); SSI scores (0.084); tuition/fee costs (0.077); ABHE Bible Content Test scores (0.066); and General Education test scores (0.055). These weights somewhat fit the Spellings Commission’s (U.S. Department of Education, 2006) perspective that accreditation values inputs and processes (costs, faculty degrees, financial factor) over outcomes (e.g., test scores), but the team evaluators’ emphasis on retention and completion (metrics favored by the Commission) suggests that accreditation participants’ perspective has greater complexity and more nuance than the Commission acknowledged. Additionally, the team evaluators’ emphasis on costs can be seen as sharing the Commission’s concern with affordability.

Metric cutpoints based on expert responses did not correspond neatly with normative cutpoints based on data from ABHE institutions, although the differences were not statistically significant (so Null Hypothesis 3 was not rejected). Across metrics, the index bands had the
Figure 9. Multivariate index of institutional health showing category ranges on a unit scale.

following ranges on a unit (1-point) scale: At risk, below 0.4494; Challenged, 0.4494-0.5550; Effective, 0.5551-0.6491; Excelling, 0.6492 and above. Figure 9 depicts the bands on the unit scale, showing their relative ranges. Interestingly, the Effective range is slightly smaller than the Challenged range, both of which are quite narrow compared to the entire unit scale (about 10%).

Objective Two: Metric validation. The second part of Phase One addressed metric validity. To consider the metric’s validity, both the literature/current accreditation practice and review by independent experts were used. The metrics that evaluators included have strong support in the literature (suggesting support for Hypothesis 4, which connects validity to the literature). But the evaluators did not order metric weights as Hypothesis 2 predicted from the
literature, leaving literature-based validation only partially supported. Responses from
independent experts, while providing limited support for the index, raised sufficient questions
about the index’s component metrics, metric weights and index cutpoints to fail to reject Null
Hypothesis 5, meaning the index is not considered to be validated by the independent experts.
Overall, then, this study failed to validate the index.

In terms of the accountability and improvement paradigms for assessment (Ewell, 2008),
ABHE team evaluators seemed not to hold strongly to one or the other when it came to metric
inclusion. Some metrics that were included and weighted more heavily lean more toward
accountability (standardized tests, costs), while others seem more improvement oriented (SSI
scores, faculty degrees, financial factor) – and some, especially retention and completion rates,
may fit both paradigms. These results suggest that accreditation may not be cleanly aligned with
improvement or accountability, but may value both, despite a tendency in the literature to take
one side or the other (Ewell, 2008; Graham et al., 1995; Metz, 2011; Trow, 1996).

Consideration of metric cutpoints (Table 12) adds to the complexity. For example, results
set 65% of students improving their scores on standardized tests (comparing their freshman and
senior scores) to be Effective, which seems a very modest threshold (if 80% of students show
improvement, the institution would fall into the Excelling category). Similarly, a completion rate
of 45% was held to be Effective. Such cutpoints can be considered fairly generous, suggesting a
certain amount of discomfort with using assessment results for accountability, with its
transparent publicity. A similar perspective was revealed in validation efforts, with one
independent expert in particular questioning the validity of these metrics for accountability.
Results in Phase One, taken as a whole, affirm the literature’s sense that there is a tension
between improvement and accountability assessment paradigms, even as they blend and overlap in actual systems (Banta, 2007; Ewell, 2008; Kuh & Ewell, 2010; Trow, 1996).

**Objective Three: Accountability template.** In Phase Two, internal and external institutional stakeholders were surveyed to determine what information they believed should be included in an online accountability template in order for them to evaluate institutional health (defined as mission fulfillment). Stakeholder groups supported inclusion of all metrics presented to them, supporting Hypothesis 6 and differing from the ABHE evaluators from Phase One. On the one hand, this reported desire for information supports the Spellings Commission’s emphasis on “consumer information” (U.S. Department of Education, 2006) and aligns well with the position that trust by consumers should be supported by actions to keep consumers informed (Graham et al., 1995). On the other hand, Phase Three’s satisfaction levels (only moderate) raise questions about how easy it will be to meet stakeholders’ desire for information.

Contrary to Hypothesis 7, a MANOVA test for differences in importance among stakeholder groups was not significant (failing to reject Null Hypothesis 7). Thus, the online accountability template was developed without different “gateways” for different stakeholder groups. Here, a small set of metrics was presented, so results are limited. But the lack of significant differences in the importance different stakeholder groups place on sources of information calls into question the need for a system allowing users to customize college rankings to emphasize their individual preferred metrics, something the Spellings Commission report (U.S. Department of Education, 2006) and, especially, two of the Commission’s issue papers (C. Miller, n.d.; C. Miller & Malandra, n.d.) called for. Instead, data were organized by category: cost, student learning/academics, student success, student experience.
It is further instructive to compare the metric orderings between stakeholders (in Phase Two) and ABHE team evaluators (in Phase One); see Figure 10. Stakeholders’ emphasis on cost-related metrics seems to confirm a view, articulated by critics of the Spellings Commission report, that public, transparent accountability ties closely to consumerism (American Association of University Professors, 2007; Ewell, 2008; Padro, 2007). But it is not so simple, since stakeholders also place greater importance on outcome tests than do team evaluators. It is unclear if team evaluators share concerns expressed in the literature that such tests have questionable utility as meaningful measures of learning (Banta & Pike, 2007; Ennis, 2008; Hamilton & Banta, 2008) or that they ignore higher-education diversity (American Association of University Professors, 2007; Brittingham, 2008; M. A. Miller, 2007; Shavelson, 2007; Shulman, 2007), or if (as the Spellings Commission argued) accreditation tends to focus on inputs rather than outcomes (U.S. Department of Education, 2006). Interestingly, stakeholders supported inclusion of faculty evaluation of student work, something with support in the literature (Banta, 2007; Basken, 2008; Eaton, 2008; Peter D. Hart Research Associates, 2008; Schneider, 2009), while evaluators did not. In any case, these results support the literature’s position that there are real differences between the consumer and the accreditation views of meaningful accountability.

**Objective Four: Stakeholder understanding and satisfaction.** Phase Three returned to Phase Two’s stakeholder groups, asking them to access four mock colleges’ profiles in the online accountability template, then use the profiles to respond to interpretive questions and satisfaction scales. As Hypothesis 8 predicted, there were significant differences among stakeholder groups’ accuracy in interpreting the profiles, with donors less accurate than parents and staff. Groups’ satisfaction with the profiles was tepid: while all groups had mean template satisfaction scores greater than 4, supporting Hypothesis 9, four groups’ satisfaction was not significantly different.
Figure 10. Metric ordering for Stakeholders (Phase Two) and ABHE evaluators (Phase One), with higher number corresponding to greater importance.

from 4 and only one group’s satisfaction was above 5. Controlled for accuracy, satisfaction did not differ significantly among groups.

This study’s most important result was the lack of strong accuracy in stakeholders’ ability to interpret the profile data. Four of ten metrics had average accuracy (across groups) lower than 50%, and all but one metric (tuition, which had 100% accuracy) had accuracy below 77%. In terms of Metz’s (2011) framework (see Figure 3), the problem arises in the transfer of meaningful information from Agent A (the college) through Manner D (the online profile) to Body C (stakeholder groups) and Agents I (stakeholders): from this study, it is unclear if the lack
of understanding is a function of the manner (online profiles) or is related to stakeholders’ lack of relevant knowledge that would enable them to interpret the profiles accurately.

The literature includes discussions of what methods and metrics to use in measuring student learning in order to present results to stakeholders (e.g., Adler-Kassner & Harrington, 2010; American Association of University Professors, 2007; Banta, 2007; Banta & Pike, 2007; Eaton, 2008; Graff & Birkenstein, 2008; Hersh, 2007; Lingenfelter, 2007; Liu, 2010; Peter D. Hart Research Associates, 2008; Schneider, 2009; Shavelson, 2007). But from these results, it appears a prior question needs to be addressed, that of helping stakeholders interpret accountability information; this supports NILOA’s call for explanation as part of institutions’ transparency efforts (Jankowski & Makela, 2010).

**Limitations**

Limited scope and limited scale present this study’s most significant limitations. Scope was limited by design to ABHE institutions, but low participation by institutions and their stakeholders raises questions as to what extent these results represent even the biblical higher-education sector.

Phase One was limited by a small respondent pool and reliance on closed-ended survey items. Numerous respondents were unfamiliar with some of the metrics presented, although all metrics are common in discussions of higher-education accountability. Many of these metrics seek to capture less tangible aspects of the college experience, including student learning and student satisfaction. But even experts, such as accreditation evaluators, may not be sufficiently familiar with all metrics to speak to their usefulness for evaluating an institution’s performance. This suggests the need for working groups or some other means to enable experts in specific areas to work together to identify useful metrics. In the context of this study, unfamiliarity or
discomfort with the metrics led to missing data and, as explained, mean imputation. The resulting attenuated relationships used in the principle components analysis, leading to reduced precision in metric weighting.

Scale was limited by the small number of respondents, especially in Phase Three. This phase addressed the most important issues, stakeholder understanding and satisfaction, but those results are heavily qualified by the small sample; rather than being firm conclusions, the phase-three results are tentative statements of concern about online accountability systems.

All phases (except validation by independent experts) were limited by the exclusive use of closed survey items, without comment opportunities. This allowed for simpler data analysis for purposes of this study, but limited the insights into why specific results were obtained. For example, comments in Phase Three might have provided better explanations of respondents’ moderate satisfaction.

**Implications and Opportunities**

From this study arise three implications, or action opportunities, as well as two clear future research needs.

**Action opportunities.** First, institutions have an opportunity to respond to stakeholders’ desire for information. Stakeholders are interested in knowing what institutions are doing, how well they are functioning. As institutions respond with transparency, they can build trust with their stakeholders, collapsing the conflict Trow (1996) sees between public accountability and trust by using them to reinforce each other.

Second, there is an opportunity for accrediting agencies to explore ways to translate their perspective on institutions for non-experts. The multivariate index designed in this study was not validated and seems a problematic way to accomplish such translation: to satisfy experts, it
would need either to be so detailed as to be unwieldy or so minimal as to be an ineffective measure of institutional health. But working within specialized accrediting agencies does enable use of more meaningful measures of learning in common areas (in this study’s example, biblical studies) and benchmarking on some instruments, both of which have support (Adler-Kassner & Harrington, 2010; Banta, 2007; Basken, 2008; Eaton, 2008; Ewell, 2011; Peterson, 2009; Schneider, 2009; Seybert, 2006). Accrediting agencies have here a methodology that can be used to explore opportunities to serve their members and their members’ stakeholders more effectively by better assuring quality.

Third, there is room for research centers, such as NILOA, to explore stakeholder understanding of and satisfaction with various initiatives. While systems such as VSA may do internal studies, third-party centers are better positioned to provide objective reviews of how stakeholders use and value transparency systems. Along with NILOA’s framework, the Degree Qualifications Profile (DQP) provides a different way of framing student learning. If stakeholder needs are better met through such structures, research centers can advance higher-education accountability by demonstrating this objectively, then assisting institutions in adopting such frameworks. Both the NILOA framework and the DQP are being promoted to institutions, but without empirical support for their usefulness to stakeholders.

**Research needs.** Three clear areas in need of additional research arise from this study. First, there is no published research on stakeholder accuracy in interpreting information in the existing transparency initiatives, such as the VSA. This study’s small scale suggests that stakeholder interpretation of information may not be very accurate, but institutions are expending resources to maintain profiles in these systems. Particularly in light of state mandates to
participate, it is important to determine the systems’ usefulness to the stakeholders they ostensibly serve.

Second, this study’s coverage of stakeholder satisfaction with online systems should be studied and expanded to address more than only “institutional health” as defined here. An easy example here is the lack of group-specific information displays: while VSA’s College Portraits seem designed to include this, and while NILOA has called for information tailored to specific audiences (Jankowski & Makela, 2010), there is no research into how different groups view and value the profiles.

Third, this study’s nearly exclusive use of quantitative, closed items created gaps in understanding. Qualitative approaches could usefully illuminate each phase. More understanding is needed of experts’ discomfort with some metrics, and possibly with the whole concept of an index of institutional health. In phases two and three, qualitative research could powerfully inform accountability discussions by clarifying what information stakeholders want and why they want it, as well as why they report lower than expected satisfaction with that data when they see it.

**Conclusion**

In January 2015, it does not appear that accountability pressures on higher-education institutions will disappear. In fact, with the White House’s announced intentions to develop a national rating system for colleges (Field, 2014; Mangan & Supiano, 2014), regulatory pressure may be increasing. Para-college organizations such as the VSA continue their efforts in this area, sometimes in response to the federal government’s actions (Will, 2015). This regulatory turn toward public systems may in turn increase accountability pressure from constituents (or “consumers”). In such a context, this study’s exploration of internal and external stakeholder
perceptions provides tentative answers to some questions, while it raises some new questions as well. Accrediting agencies also face scrutiny – in December 2014, the National Advisory Committee on Institutional Quality and Integrity (NACIQI) discussed, among other things, a proposal to dismantle the regional system that currently defines the most prestigious accrediting agencies (Kelderman, 2014).

This study’s efforts to develop a multivariate index of institutional health with performance bands provides relevant insights for systems such as the White House’s draft framework, released in December 2014 (see http://www2.ed.gov/documents/college-affordability/framework-invitation-comment.pdf) (Field, 2014). It has been noted that achieving acceptable consensus around a rating system likely would require either greater complexity or minimalism. The White House’s December 2014 draft framework shows an attempt to avoid controversy by simply not including measures of student learning or student satisfaction, aligning with the minimalist option. But Phase Two concluded that stakeholders are information-hungry. This makes it unclear that a system such as the White House has proposed will be very meaningful to stakeholders, including the consumer and regulatory stakeholders that the White House, in particular, aims to serve (Field, 2014; Supiano, 2014; Thomason, 2014). A minimal system may avoid controversy, but also avoid usefulness.

Most importantly, this study raises serious questions about stakeholders’ ability to interpret accountability data with consistent high accuracy. This poses a problem for all existing transparency initiatives, as well as those under development: identifying valid and meaningful data to present accomplishes little if the intended audiences cannot use that data effectively. Essentially this is not accountability, since those in positions to hold institutions accountable – consumers, donors, other stakeholders – still de facto lack the understanding to do so. A prior
question has been overlooked: how can we educate stakeholders to interpret accountability information in order to empower them to exercise their accountability powers? This presents both a challenge and an opportunity, an opportunity for additional research as well as for meaningful action by institutions and accrediting agencies, as well as by regulatory and other stakeholders.
References


Lantz, B. (2013). The impact of sample non-normality on ANOVA and alternative methods.  
*British Journal of Mathematical & Statistical Psychology, 66*(2), 224–244.  

Retrieved from  


doi:10.3102/0013189X020008015


Brookings Institution. Retrieved from  

doi:10.3102/0013189X11424314


Peters, R. (1994). Some snarks are boojums: Accountability and the end(s) of higher education. 
*Change*, 26(6), 16–23.


Trusted information on returning to college: College Choices for Adults. (2010, August 27). *PR Newswire*. Boulder, CO.


Appendix A: Screenshots for Survey of Expert Evaluators for Phase One
Below is a list of possible metrics to include in a weighted accountability index. For each metric, please indicate whether or not you think it should be included in the index.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Yes, include in index</th>
<th>No, do not include in index</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Education test freshman and senior scores (also include ETS Proficiency Profile or ACT COMP)</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td>Completion rate</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Retention rate</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Noel-Lehto Student Satisfaction Inventory (NSI) scores on reading items</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>National Survey of Student Engagement (NSSE) scores on five benchmarks</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Aggregate faculty evaluations of senior-level work based on ABHE metrics</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>ABET Score Content Test freshman and senior scores</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Percentage of full-time faculty with doctoral degrees</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Full-time tuition and required fees compared with ABHE average</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>ABET total weighted factor (weighted finance ratio)</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Room and board charges compared with ABHE average</td>
<td>☑</td>
<td>☑</td>
</tr>
</tbody>
</table>

Below are the metrics you chose for inclusion in a multivariate index. For each metric, please rate its relative importance.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Not important at all</th>
<th>Not very important</th>
<th>Somewhat unimportant</th>
<th>Neutral</th>
<th>Somewhat important</th>
<th>Important</th>
<th>Very important</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABET Score Content Test freshman and senior scores</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>General Education test freshman and senior scores (also include ETS Proficiency Profile or ACT COMP)</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Aggregate faculty evaluations of senior-level work based on ABHE metrics</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>National Survey of Student Engagement (NSSE) scores on five benchmarks</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Noel-Lehto Student Satisfaction Inventory (NSI) scores on reading items</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Full-time tuition and required fees compared with ABHE average</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Completion rate</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Percentage of full-time faculty with doctoral degrees</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>ABET total weighted factor (weighted finance ratio)</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
</tbody>
</table>
In order to create "bands" or ranges of institutional health on the index, cut points are needed. Four institutional bands or categories are planned (see diagram below):

- **At-risk institutions** have problems that inhibit their ability to comply substantially with ASHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged institutions** work to maintain substantial compliance with ASHE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- **Effective institutions** substantially meet expectations in ASHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Excelling institutions** surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

On the following pages, you will be asked a series of questions to help establish boundaries for these four categories. If you do not feel comfortable identifying boundaries for a particular area (e.g., academic, testing, financial ratios, student satisfaction, etc.), you may indicate this on an "opt-out" question on each page.
The questions below ask you to suggest category boundaries related to general-education testing. The four categories (At-risk, Challenged, Effective, and Exciting) were presented above. The descriptions are repeated here for your convenience:

- At-risk institutions have problems that inhibit their ability to comply substantially with ASHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are addressed.
- Challenged institutions work to maintain substantial compliance with ASHE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- Effective institutions substantially meet expectations in ASHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- Exciting institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

Questions assume use of one of two tests: the Collegiate Assessment of Academic Proficiency (CAAP) or the Proficiency Profile. In answering the questions below, please consider the above descriptions of institutional categories and the following normative/benchmark information:

- On the CAAP in Fall 2011, the average score for freshmen was 56.06. The average score for seniors was 66.82. These averages include all 5 test domains (Reading, Writing Skills, Mathematics, Science, Critical Thinking), each of which is scored on a scale ranging from 40 to 80.
- On the Proficiency Profile from 1/2008 to 4/2011, the average score for freshmen at accredited institutions was 434.41. The average score for seniors was 436.86. Scores range from 400 to 600.

Consider the above definition of an Exciting institution and the above norm/ benchmarking information. For an institution to be classified as Exciting, what is the minimum percentage of graduating students whose senior score must be higher than their freshman score on one of these tests? (Note that this minimum separates Exciting institutions from Effective institutions.)

- 10%
- 15%
- 20%
- 25%
- 30%
- 35%
- 40%
- 45%

Consider the above definition of an Effective institution and the above norm/ benchmarking information. For an institution to be classified as Effective, what is the minimum percentage of graduating students whose senior score must be higher than their freshman score on one of these tests? (Note that this minimum separates Effective institutions from Challenged institutions.)

- 10%
- 15%
- 20%
- 25%
- 30%
- 35%
- 40%
- 45%

Consider the above definition of a Challenged institution and the above norm/ benchmarking information. For an institution to be classified as Challenged, what is the minimum percentage of graduating students whose senior score must be higher than their freshman score on one of these tests? (Note that this minimum separates Challenged institutions from At-risk institutions.)

- 10%
- 15%
- 20%
- 25%
- 30%
- 35%
- 40%
- 45%

If you left the above question blank because you do not feel comfortable suggesting boundaries for general-education testing, please indicate so here. Otherwise, please leave this question unanswered.

- I am not comfortable suggesting boundaries in this category.
The questions below ask you to suggest category boundaries related to biblical knowledge testing. The four categories (All-risk institutions, Challenged institutions, Effective institutions, Excelling institutions) were presented above. The descriptions are repeated here for your convenience.

- **All-risk** institutions have problems that inhibit their ability to comply substantially with ABHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged** institutions work to maintain substantial compliance with ABHE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- **Effective** institutions substantially meet expectations in ABHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Excelling** institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic, educational, student satisfaction, operational integrity, finances).

Questions assume use of the ABHE Bible Content Exam.

Consider the above definition of an **Excelling** institution and the above norm/benchmarking information. For an institution to be classified as Excelling, what is the minimum percentage of graduating students whose senior score must be higher than their freshman score on the ABHE Bible Content Exam? (Note that this minimum separates Excelling institutions from Effective institutions.)

<table>
<thead>
<tr>
<th></th>
<th>75%</th>
<th>70%</th>
<th>65%</th>
<th>60%</th>
<th>55%</th>
<th>50%</th>
<th>45%</th>
<th>40%</th>
<th>35%</th>
<th>30%</th>
<th>25%</th>
<th>20%</th>
<th>15%</th>
<th>10%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consider the above definition of an **Effective** institution and the above norm/benchmarking information. For an institution to be classified as Effective, what is the minimum percentage of graduating students whose senior score must be higher than their freshman score on the ABHE Bible Content Exam? (Note that this minimum separates Effective institutions from Challenged institutions.)

<table>
<thead>
<tr>
<th></th>
<th>75%</th>
<th>70%</th>
<th>65%</th>
<th>60%</th>
<th>55%</th>
<th>50%</th>
<th>45%</th>
<th>40%</th>
<th>35%</th>
<th>30%</th>
<th>25%</th>
<th>20%</th>
<th>15%</th>
<th>10%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consider the above definition of an **Challenged** institution and the above norm/benchmarking information. For an institution to be classified as Challenged, what is the minimum percentage of graduating students whose senior score must be higher than their freshman score on the ABHE Bible Content Exam? (Note that this minimum separates Challenged institutions from All-risk institutions.)

<table>
<thead>
<tr>
<th></th>
<th>75%</th>
<th>70%</th>
<th>65%</th>
<th>60%</th>
<th>55%</th>
<th>50%</th>
<th>45%</th>
<th>40%</th>
<th>35%</th>
<th>30%</th>
<th>25%</th>
<th>20%</th>
<th>15%</th>
<th>10%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you left the above question blank because you do not feel comfortable suggesting boundaries for biblical knowledge testing, please indicate so here. Otherwise, please leave this question unanswered.

I am not comfortable suggesting boundaries in this category.
The questions below ask you to suggest category boundaries related to faculty evaluations of senior-level students' work. The four categories (Al-risk institutions, Challenged institutions, Effective institutions, Exceling institutions) were presented above. The descriptions are repeated here for your convenience.

- **At-risk** institutions have problems that inhibit their ability to comply substantially with ABHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged** institutions work to maintain substantial compliance with ABHE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- **Effective** institutions substantially meet expectations in ABHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They maintain a sound financial standing.
- **Exceling** institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

A 2010 ABHE project presented at the ABHE annual meeting identified 8 likely outcomes for ABHE institutions and developed rubrics for each outcome. Please view the rubrics by clicking here (notes on the bottom of the spreadsheet show all 6 rubrics).

**Consider the above definition of an Exceling institution and the above norm/benchmark information. For an institution to be classified as Exceling, what is the minimum percentage of graduating students whose work should be evaluated by faculty as Proficient or Exemplary on the rubric? (Note: This minimum separates Exceling institutions from Effective institutions.)**

<table>
<thead>
<tr>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>30%</td>
<td>35%</td>
<td>40%</td>
<td>45%</td>
<td>50%</td>
<td>55%</td>
<td>60%</td>
<td>65%</td>
<td>70%</td>
</tr>
</tbody>
</table>

**Consider the above definition of an Effective institution and the above norm/benchmark information. For an institution to be classified as Effective, what is the minimum percentage of graduating students whose work should be evaluated by faculty as Proficient or Exemplary on the rubric? (Note: This minimum separates Effective institutions from Challenged institutions.)**

<table>
<thead>
<tr>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>30%</td>
<td>35%</td>
<td>40%</td>
<td>45%</td>
<td>50%</td>
<td>55%</td>
<td>60%</td>
<td>65%</td>
<td>70%</td>
</tr>
</tbody>
</table>

**Consider the above definition of a Challenged institution and the above norm/benchmark information. For an institution to be classified as Challenged, what is the minimum percentage of graduating students whose work should be evaluated by faculty as Proficient or Exemplary on the rubric? (Note: This minimum separates Challenged institutions from At-risk institutions.)**

<table>
<thead>
<tr>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>30%</td>
<td>35%</td>
<td>40%</td>
<td>45%</td>
<td>50%</td>
<td>55%</td>
<td>60%</td>
<td>65%</td>
<td>70%</td>
</tr>
</tbody>
</table>

If you left the above question blank because you do not feel comfortable suggesting boundaries for faculty evaluation of senior work with rubrics, please indicate so here. Otherwise, please leave the question unanswered.

**I am not comfortable suggesting boundaries in this category.**

---

**Survey Planned By:** Datashield
The questions below ask you to suggest category boundaries related to institutions’ level of student engagement. The four categories (At-risk institutions, Challenged institutions, Effective institutions, Excelling institutions) were presented above. The descriptions are repeated here for your convenience:

- At-risk institutions have problems that inhibit their ability to comply substantially with ASHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- Challenged institutions work to maintain substantial compliance with ASHE criteria, with particular challenges in one or two areas (academic/educational; student satisfaction; operational integrity; finances).
- Effective institutions substantially meet expectations in ASHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- Excelling institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational; student satisfaction; operational integrity; finances).

NSSE, the National Survey of Student Engagement, annually collects information at hundreds of four-year colleges and universities about student participation in programs and activities that institutions provide for their learning and personal development. The results provide an estimate of how undergraduates spend their time and what they gain from attending college, according to the NSSE website (http://nsse.iub.edu). From these results five benchmarks are developed; here are 2011 average scores for seniors on the NSSE benchmarks (each is on a 0-100 scale). (Click the benchmark name for NSSE’s definition.)

- Academic Challenge: 57.4
- Active/collegiate Learning: 51.3
- Student/Faculty Interaction: 42.6
- Envisioning Educational Experience: 40.4
- Supportive Campus Environment: 59.1

Consider the definition of an Excelling institution and the NSSE benchmark averages above. For an institution to be classified as Excelling, what is the minimum average score an institution’s seniors should have on the 0-100 scale? Please provide a score for each benchmark area. (Note that this minimum separates Excelling and Effective institutions.)

- Academic Challenge
- Active/collegiate Learning
- Student/Faculty Interaction
- Envisioning Educational Experience
- Supportive Campus Environment

Consider the definition of an Excelling institution and the NSSE benchmark averages above. For an institution to be classified as Excelling, what is the minimum average score an institution’s seniors should have on the 0-100 scale? Please provide a score for each benchmark area. (Note that this minimum separates Excelling and Effective institutions.)

- Academic Challenge
- Active/collegiate Learning
- Student/Faculty Interaction
- Envisioning Educational Experience
- Supportive Campus Environment

Consider the definition of an Effective institution and the NSSE benchmark averages above. For an institution to be classified as Effective, what is the minimum average score an institution’s seniors should have on the 0-100 scale? Please provide a score for each benchmark area. (Note that this minimum separates Effective and Challenged institutions.)

- Academic Challenge
- Active/collegiate Learning
- Student/Faculty Interaction
- Envisioning Educational Experience
- Supportive Campus Environment

If you left the above question blank because you do not feel comfortable suggesting boundaries for student engagement assessment, please indicate so here. Otherwise, please leave this question unanswered. (I am not comfortable suggesting boundaries in this category.)
The questions below ask you to suggest category boundaries related to institutions’ scores on student satisfaction scales. The four categories (At-risk institutions, Challenged institutions, Effective institutions, Exciting institutions) were presented above. The description are repeated here for your convenient reference:

- **At-risk institutions** have problems that inhibit their ability to comply substantially with the criteria. These institutions are generally small, under-funded, have low faculty retention, and have low faculty salaries.
- **Challenged institutions** score too low to meet the educational standards set by their peers. They have problems with faculty retention, faculty salaries, student achievement, and research and development activities.
- **Effective institutions** substantially meet expectations in all criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Exciting institutions** also meet the expectations for compliance. They provide members of a high-quality institution of higher education in multiple areas (academic, educational, student satisfaction, operational integrity, finances).

The Student Satisfaction Inventory produced from the five levels, “measures student satisfaction and priorities, showing how satisfied students are as well as what issues are reported to their.” According to the five levels above, the Student Satisfaction Inventory for the University of Cincinnati includes the following questions, each of which was scored on a 1-7 scale. The central point at the level for each scale is 4. (Each scale name for Web-CAT’s description.)

### Academic-Administrative Effectiveness
- **Campus Climate**
- **Campus Life**
- **Campus Students**
- **Instructional Effectiveness**
- **Operation and Financial Administration**
- **Regulation Effectiveness**
- **Safety and Security**
- **Student Concerns**

Consider the definition of an **Academic** institution and the student scores above. For an institution to be classified as **Exciting**, what is the **minimum** average score an institution should have on the 7-point scale? Please provide a score for each scale. (Note that the minimum separates Exciting and Effective institutions.)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campus Climate</td>
<td>4</td>
</tr>
<tr>
<td>Campus Life</td>
<td>4</td>
</tr>
<tr>
<td>Campus Students</td>
<td>4</td>
</tr>
<tr>
<td>Instructional Effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>Operation and Financial Administration</td>
<td>4</td>
</tr>
<tr>
<td>Regulation Effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>Safety and Security</td>
<td>4</td>
</tr>
<tr>
<td>Student Concerns</td>
<td>4</td>
</tr>
</tbody>
</table>

### Academic-Administrative Effectiveness

Consider the definition of an **Effective** institution and the student scores above. For an institution to be classified as **Effective**, what is the **minimum** average score an institution should have on the 7-point scale? Please provide a score for each scale. (Note that the minimum separates Effective and Challenged institutions.)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campus Climate</td>
<td>4</td>
</tr>
<tr>
<td>Campus Life</td>
<td>4</td>
</tr>
<tr>
<td>Campus Students</td>
<td>4</td>
</tr>
<tr>
<td>Instructional Effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>Operation and Financial Administration</td>
<td>4</td>
</tr>
<tr>
<td>Regulation Effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>Safety and Security</td>
<td>4</td>
</tr>
<tr>
<td>Student Concerns</td>
<td>4</td>
</tr>
</tbody>
</table>

### Academic-Administrative Effectiveness

Consider the definition of an **Challenged** institution and the student scores above. For an institution to be classified as an **Effective**, what is the **minimum** average score an institution should have on the 7-point scale? Please provide a score for each scale. (Note that the minimum separates Challenged and At-risk institutions.)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campus Climate</td>
<td>4</td>
</tr>
<tr>
<td>Campus Life</td>
<td>4</td>
</tr>
<tr>
<td>Campus Students</td>
<td>4</td>
</tr>
<tr>
<td>Instructional Effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>Operation and Financial Administration</td>
<td>4</td>
</tr>
<tr>
<td>Regulation Effectiveness</td>
<td>4</td>
</tr>
<tr>
<td>Safety and Security</td>
<td>4</td>
</tr>
<tr>
<td>Student Concerns</td>
<td>4</td>
</tr>
</tbody>
</table>

If you feel that the levels listed do not provide a comfortable suggesting boundaries for student-satisfaction assessment, please indicate so. Otherwise, please leave this question unanswered.

© I do not contribute suggesting boundaries in this category.
The questions on this page and the next ask you to suggest category boundaries related to institutional tuition/fellowship and room/board. The four categories (All-risk Institutions, Challenged institutions, Effective institutions, Exciting institutions) were presented above. The descriptions are repeated here for your convenience and reference.

- **All-risk institutions** have problems that inhibit their ability to comply substantially with ABHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged institutions** work to maintain substantial compliance with ABHE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- **Effective institutions** substantially meet expectations in ABHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Exciting institutions** surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

Fulltime tuition means published undergraduate tuition; fees means required undergraduate fees. This information is supplied annually by ABHE member institutions to ABHE. Here are average amounts from ABHE’s 2006-2010 Statistical report compiled from the 2010 annual report:

- Average fulltime tuition: $9,035.10
- Average fees: $502.10
- Average total tuition/fees: $9,537.20

Tuition/fees and room/board can be seen as best if they are moderate; if they are too low, an institution’s financial health may be at risk, if they are too high, an institution’s ability to recruit and retain students may be at risk. Consequently, the questions below ask you to identify how far from the ABHE average an institution should be—that is, assume that closer to average is usually better, with a large difference above or below the average as dangerous.

Consider the definition of an *Exciting* institution and the ABHE averages above. For an institution to be classified as *Exciting*, what is the minimum percentage above or below the ABHE average an institution’s tuition/fees should be? (Note that this minimum separates Exciting and Effective institutions.)

| 100 | 70 | 45 | 20 |
| 95 | 65 | 40 | 15 |
| 85 | 55 | 30 | 5 |
| 75 | 50 | 25 | 0 |

Consider the definition of an *Effective* institution and the ABHE averages above. For an institution to be classified as *Effective*, what is the minimum percentage above or below the ABHE average an institution’s tuition/fees should be? (Note that this minimum separates Effective and Challenged institutions.)

| 100 | 70 | 45 | 20 |
| 95 | 65 | 40 | 15 |
| 85 | 55 | 30 | 5 |
| 75 | 50 | 25 | 0 |

Consider the definition of an *Challenged* institution and the ABHE averages above. For an institution to be classified as *Challenged*, what is the minimum percentage above or below the ABHE average an institution’s tuition/fees should be? (Note that this minimum separates Challenged and All-risk institutions.)

| 100 | 70 | 45 | 20 |
| 95 | 65 | 40 | 15 |
| 85 | 55 | 30 | 5 |
| 75 | 50 | 25 | 0 |

If you left the above questions blank because you do not feel comfortable suggesting boundaries for tuition/fees, please indicate so here. Otherwise, please leave this question unanswered.

- I am not comfortable suggesting boundaries in this category.

Survey Powered By Askiq
This page contains questions about roomboard amounts. The descriptions of institutional categories are repeated here for your convenient reference:

- **At-risk** institutions have problems that inhibit their ability to comply substantially with ABHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged** institutions work to maintain substantial compliance with ABHE criteria, with particular challenges in one of two areas: academic/educational, student satisfaction, operational integrity, finances.
- **Effective** institutions substantially meet expectations in ABHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Excelling** institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas: academic, educational, student satisfaction, operational integrity, finances.

Roomboard means published undergraduate room and board. This information is supplied annually by ABHE member institutions to ABHE. The average amount in ABHE's 2009-2010 Statistical Report compiled from the 2010 annual report was $4,937.80.

Tuition/fees and roomboard can be seen as best if they are moderate, if they are too low, an institution's financial health may be at risk. If they are too high, an institution's ability to recruit and retain students may be at risk. Consequently, the questions below ask you to identify how far from the ABHE average an institution should be - that is, assume that closer to average is usually better, with a large difference above or below the average as dangerous.

Consider the definition of an **Excelling** institution and the ABHE averages above. For an institution to be classified as **Excelling**, what is the maximum percentage above or below the ABHE average an institution's roomboard should be? (Note that this minimum separates Excelling and Effective institutions.)

| 100 | 70 | 45 | 20 |
| 95 | 65 | 40 | 15 |
| 90 | 60 | 35 | 10 |
| 85 | 55 | 30 | 5 |
| 80 | 50 | 25 | 0 |
| 75 |   |    |    |

Consider the definition of an **Effective** institution and the ABHE averages above. For an institution to be classified as **Effective**, what is the maximum percentage above or below the ABHE average an institution's roomboard should be? (Note that this minimum separates Effective and Challenged institutions.)

| 100 | 70 | 45 | 20 |
| 95 | 65 | 40 | 15 |
| 90 | 60 | 35 | 10 |
| 85 | 55 | 30 | 5 |
| 80 | 50 | 25 | 0 |
| 75 |   |    |    |

Consider the definition of a **Challenged** institution and the ABHE averages above. For an institution to be classified as **Challenged**, what is the maximum percentage above or below the ABHE average an institution's roomboard should be? (Note that this minimum separates Challenged and At-risk institutions.)

| 100 | 70 | 45 | 20 |
| 95 | 65 | 40 | 15 |
| 90 | 60 | 35 | 10 |
| 85 | 55 | 30 | 5 |
| 80 | 50 | 25 | 0 |
| 75 |   |    |    |

If you left the above questions blank because you do not feel comfortable suggesting boundaries for roomboard, please indicate so here. Otherwise, please leave this question unanswered. (I am not comfortable suggesting boundaries in this category)
The questions below ask you to suggest category boundaries related to an institution’s total weighted factor, a composite of financial ratios. For your reference, the category descriptions are repeated here.

- **AI-risk institutions** have problems that inhibit their ability to comply substantially with ABHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged** institutions work to maintain substantial compliance with ABHE criteria, with particular challenges in one or two areas (academic educational, student satisfaction, operational integrity, finances).
- **Effective** institutions substantially meet expectations in ABHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Excelling** institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic educational, student satisfaction, operational integrity, finances).

The total weighted factor is collected annually by ABHE as part of the institutional annual report. The factor ranges from -1.0 to +1.0. According to the U.S. Department of Education’s Federal Student Aid Handbook (vol. 2, ch. 17), the total weighted factor may be interpreted by the following standards:

- **-1.0 to -0.6**: Institution is “financially responsible; no additional oversight required.”
- **-0.6 to 0.0**: Institution is “financially responsible; additional oversight recommended.”
- **0.0 to +0.4**: Institution is “financially responsible; additional financial aid requirements are reviewed.”
- **+0.4 to +0.9**: Institution is “financially responsible; additional requirements will be applied.”
- **+0.9 to +1.0**: Institution is “financially responsible; additional requirements will be applied.”

### Consider the definition of an **Excelling** institution and the information on interpreting the total weighted factor above. For an institution to be classified as **Excelling**, what is the **minimum** total weighted factor score it should achieve? (Note that this minimum separates Excelling and Effective institutions.)

<table>
<thead>
<tr>
<th>Score</th>
<th>0.0</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.75</td>
<td>1.50</td>
<td>1.25</td>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Consider the definition of an **Effective** institution and the information on interpreting the total weighted factor above. For an institution to be classified as **Effective**, what is the **maximum** total weighted factor score it should achieve? (Note that this minimum separates Effective and Challenged institutions.)

<table>
<thead>
<tr>
<th>Score</th>
<th>0.0</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.75</td>
<td>1.50</td>
<td>1.25</td>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Consider the definition of an **Challenged** institution and the information on interpreting the total weighted factor above. For an institution to be classified as **Challenged**, what is the **minimum** total weighted factor score it should achieve? (Note that this minimum separates Challenged and AI-risk institutions.)

<table>
<thead>
<tr>
<th>Score</th>
<th>0.0</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.75</td>
<td>1.50</td>
<td>1.25</td>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

If you left the above questions blank because you do not feel comfortable suggesting boundaries for the total weighted factor, please indicate so here. Otherwise, please leave this question unanswered.

- I am not comfortable suggesting boundaries in this category
The questions below ask you to suggest category boundaries related to an institution’s faculty credentials. For your reference, the category descriptions are repeated here:

- **All-risk** institutions have problems that inhibit their ability to comply substantially with ABHE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged** institutions work to maintain substantial compliance with ABHE criteria, with particular challenges in one of two areas (academic/educational, student satisfaction, operational/financial).
- **Effective** institutions substantially meet expectations in ABHE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Excelling** institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

The percentage of full-time faculty with doctoral degrees is supplied annually by ABHE member institutions to ABHE. ABHE’s 2009-2010 Statistical Report compiled from the 2010 annual report reported an average percentage of 50.3%. The National Center for Education Statistics reported from its most recent study of higher education faculty (2005-2006) a national percentage of 64.6% in private four-year institutions.

Consider the above definition of an **Excelling** institution and the ABHE and national benchmarks above. For an institution to be classified as **Excelling**, what is the minimum percentage of full-time faculty who should have doctorate degrees? (Note that this minimum separates Excelling institutions from Effective institutions.)

- 10%
- 20%
- 30%
- 40%
- 50%
- 60%
- 70%
- 80%
- 90%
- 100%

Consider the above definition of an **Effective** institution and the ABHE and national benchmarks above. For an institution to be classified as **Effective**, what is the minimum percentage of full-time faculty who should have doctorate degrees? (Note that this minimum separates Effective institutions from Challenged institutions.)

- 10%
- 20%
- 30%
- 40%
- 50%
- 60%
- 70%
- 80%
- 90%
- 100%

Consider the above definition of a **Challenged** institution and the ABHE and national benchmarks above. For an institution to be classified as **Challenged**, what is the minimum percentage of full-time faculty who should have doctorate degrees? (Note that this minimum separates Challenged institutions from All-risk institutions.)

- 10%
- 20%
- 30%
- 40%
- 50%
- 60%
- 70%
- 80%
- 90%
- 100%

If you left the above questions blank because you do not feel comfortable suggesting boundaries for faculty credentials, please indicate so here. Otherwise, please leave this question unanswered.

I am not comfortable suggesting boundaries in this category.
The questions on this page and the next page ask you to suggest category boundaries related to an institution’s retention and graduation rates. For your reference, the institutional category descriptions are repeated here:

- **Risk** institutions have problems that inhibit their ability to comply substantially with AHEE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged** institutions work to maintain substantial compliance with AHEE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- **Effective** institutions substantially meet expectations in AHEE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Excelling** institutions surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

Retention rate here means retention of first-time, full-time first-year students from their first fall semester to their second fall semester. This information is supplied annually by AHEE member institutions to AHEE. AHEE’s 2005-2010 Statistical Report (compiled from the 2010 annual report) reported an average retention rate of 62.7%. More broadly, ACT in fall 2011 reported a rate of 66.0% for 644 4-year private institutions and a rate of 67.1% for all institutions.

Consider the definition of an **Existing** institution and the AHEE and national retention benchmarks above. For an institution to be classified as **Existing**, what is the minimum percentage its first year to second-year retention rate should be? (Note that this minimum separates Excellent and Effective institutions.)

- 100%
- 75%
- 50%
- 25%

Consider the definition of an **Effective** institution and the AHEE and national retention benchmarks above. For an institution to be classified as **Effective**, what is the minimum percentage its first year to second-year retention rate should be? (Note that this minimum separates Effective and Challenged institutions.)

- 100%
- 75%
- 50%
- 25%

Consider the definition of an **Challenged** institution and the AHEE and national retention benchmarks above. For an institution to be classified as **Challenged**, what is the minimum percentage its first year to second-year retention rate should be? (Note that this minimum separates Challenged and Risk institutions.)

- 100%
- 75%
- 50%
- 25%

If you left the above questions blank because you do not feel comfortable suggesting boundaries for retention rate, please indicate so here. Otherwise, please leave this question unanswered.

- I am not comfortable suggesting boundaries in this category.
For your reference, the institutional category descriptions are repeated here:

- **At-risk institutions** have problems that inhibit their ability to comply substantially with AHEE criteria. Their problems are serious enough that they may face either closure or loss of accreditation unless the problems are resolved.
- **Challenged institutions** work to maintain substantial compliance with AHEE criteria, with particular challenges in one or two areas (academic/educational, student satisfaction, operational integrity, finances).
- **Effective institutions** substantially meet expectations in AHEE criteria. They provide a good education to their students at a fair price, producing satisfied students and alumni. They operate with integrity. They maintain a sound financial standing.
- **Exceling institutions** surpass minimal expectations for compliance. They provide models of a high-quality institution of higher education in multiple areas (academic/educational, student satisfaction, operational integrity, finances).

Completion rates here mean the percentage of first-time, full-time students who complete their degree program within 150% of the published degree length (3 years for a 3-year degree, 5 years for a 4-year degree, etc.). This information is supplied annually by AHEE member institutions to AHEE. AHEE's 2008-2009 Statistical Report (compiled from the 2010 annual report) reported an average completion rate of 44.5%. The National Center for Education Statistics reported in 2011 a national rate of 33.8% (determined from a representative sample).

Consider the definition of an **Exceling** institution and the AHEE and national retention benchmarks above. For an institution to be classified as **Exceling**, what is the minimum percentage its completion rate should be? (Note that this minimum separates Exceling and Effective institutions):

- 90%  
- 85%  
- 80%  
- 75%  
- 70%  
- 65%  
- 60%  
- 55%  
- 50%  
- 45%  
- 40%  
- 35%  
- 30%  
- 25%  
- 20%  
- 15%  
- 10%  
- 5%  

Consider the definition of an **Effective** institution and the AHEE and national retention benchmarks above. For an institution to be classified as **Effective**, what is the minimum percentage its completion rate should be? (Note that this minimum separates Effective and Challenged institutions):

- 90%  
- 85%  
- 80%  
- 75%  
- 70%  
- 65%  
- 60%  
- 55%  
- 50%  
- 45%  
- 40%  
- 35%  
- 30%  
- 25%  
- 20%  
- 15%  
- 10%  
- 5%  

Consider the definition of an **Challenged** institution and the AHEE and national retention benchmarks above. For an institution to be classified as **Challenged**, what is the minimum percentage its completion rate should be? (Note that this minimum separates Challenged and At-risk institutions):

- 90%  
- 85%  
- 80%  
- 75%  
- 70%  
- 65%  
- 60%  
- 55%  
- 50%  
- 45%  
- 40%  
- 35%  
- 30%  
- 25%  
- 20%  
- 15%  
- 10%  
- 5%  

If you left the above question blank because you do not feel comfortable suggesting boundaries for completion rate, please indicate so here. Otherwise, please leave this question unanswered.

- I am not comfortable suggesting boundaries in this category.
Appendix B: ABHE Draft Rubrics

Below are rubrics developed during an ABHE assessment project. The project identified eight learning outcomes that were held to be appropriate for all ABHE institutions, based on input from about half of the accredited ABHE institutions. Rubrics are presented below along with their objective statements (Mort et al., 2011).

**Outcome 1:** Exhibit knowledge of the Bible and essential Christian doctrine by interpreting Scripture through proper exegetical techniques.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Able to verbalize the backgrounds and major themes of the books of the Bible and to trace the flow of redemptive history from creation to consummation, identifying key historical events</td>
<td>1 Able to verbalize some of the backgrounds and major themes of the books of the Bible and to trace most of redemptive history from creation to consummation, identifying most key historical events</td>
<td>1 Has only a vague awareness of general content in each book of the Bible and a vague understanding of the flow of redemptive history from creation to consummation; able to identify a few key historical events</td>
</tr>
<tr>
<td>2 Able to compare and contrast the key aspects of the major topics in Christian doctrine with biblical support and can explain the support for differing viewpoints.</td>
<td>2 Able to define major topics in Christian doctrine with some biblical support but does not understand the cause of differing viewpoints.</td>
<td>2 Has a familiarization with major topics in Christian doctrine but lacks the knowledge of biblical support for the doctrine or the causes for differing viewpoints.</td>
</tr>
<tr>
<td>3 Demonstrates proficiency in analyzing and interpreting the original meaning of scripture through the use of appropriate tools and exegetical skills and the ability to defend the interpretation against differing viewpoints.</td>
<td>3 Demonstrates adequate proficiency in analyzing and interpreting the original meaning of scripture through the use of appropriate tools and exegetical skills but lacks the ability to defend the interpretation against differing viewpoints.</td>
<td>3 Demonstrates a lack of proficiency in using exegetical tools and skills or the ability to defend interpretation of scripture.</td>
</tr>
</tbody>
</table>
Outcome 2: Practice spiritual disciplines necessary for a lifelong and maturing personal relationship with Jesus Christ.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstrates the knowledge of specific spiritual disciplines for Christian growth and reasons for practicing them. Can discuss personal integration of a variety of spiritual disciplines and identify evidences of growth in a personal relationship with Jesus Christ.</td>
<td>Demonstrates the knowledge of specific spiritual disciplines for Christian growth and can identify specific instances when spiritual disciplines are practiced but one or more key disciplines are not being practiced. Can identify evidences in life of having a relationship with Jesus Christ.</td>
<td>Demonstrated the knowledge of some spiritual disciplines for Christian growth and the beginning of a personal relationship with Jesus Christ. Can relate a few instances of having practiced a few specific spiritual disciplines.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidences the outworking of Christian maturity through consistent speech and action as evaluated by others and through self-assessment.</td>
<td>Evidences the outworking of Christian maturity through generally consistent speech and action as evaluated by others and through self-assessment.</td>
<td>Evidences the need for significant growth in Christian maturity in speech and action as evaluated by others and through self-assessment.</td>
</tr>
</tbody>
</table>
**Outcome 3:** Articulate a broad range of knowledge in the natural and social sciences, humanities, and the arts as part of a well-rounded Christian life based upon a biblical world view.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Can articulate how Christian faith intersects with the natural and social sciences, the humanities and the arts. Provides a thoughtful response to the issues in these fields that seem contrary to biblical revelation.</td>
<td><strong>1</strong> Has an understanding of the natural and social sciences, the humanities and the arts. Is able to explain how Christian faith intersects with these.</td>
<td><strong>1</strong> Has a basic understanding of the natural and social sciences, the humanities and the arts. Still has trouble reconciling data that conflicts with biblical revelation.</td>
</tr>
<tr>
<td><strong>2</strong> Given a contemporary problem from a randomly selected field of the arts and sciences, the student demonstrates an expert ability to bring both current research and biblical insight to bear upon that problem in suggesting a potential solution.</td>
<td><strong>2</strong> Given a contemporary problem from a randomly selected field of the arts and sciences, the student demonstrates a basic ability to bring both current research and biblical insight to bear upon that problem in suggesting a potential solution.</td>
<td><strong>2</strong> Given a contemporary problem from a randomly selected field from the arts and sciences, the student demonstrates only a scant ability to bring both current research and biblical insight to bear upon that problem in suggesting a potential solution.</td>
</tr>
<tr>
<td><strong>3</strong> Able to demonstrate a broad understanding of human knowledge. Able to talk cogently about contemporary events/customs/culture and their impact on Christian living.</td>
<td><strong>3</strong> Able to demonstrate a basic understanding of human knowledge. Able to share about contemporary events/customs/culture and their impact on Christian living.</td>
<td><strong>3</strong> Able to grapple with the &quot;life issues&quot; of human existence and in a minimal way talk about their impact on Christian living.</td>
</tr>
<tr>
<td><strong>4</strong> Actively engages a world view on the basis of its own presuppositions, expressing it in the best possible light. Proposes realistic ways to extend a world view to life and ministry. Draws conclusions, formulates a mature response and makes applications consistent with articulated world view.</td>
<td><strong>4</strong> Demonstrates sympathy to the issues that lead a world view to function as it does. Insights are helpful and generally consistent with their world view, but not fully developed.</td>
<td><strong>4</strong> Identifies issues connected with a world view, although does not always present them carefully or charitably. Proposed applications are unrealistic; too broad or too general.</td>
</tr>
</tbody>
</table>
**Outcome 4:** Use analytical and research skills in locating, evaluating, and applying information for life and ministry.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Demonstrates the ability to locate, evaluate, and organize data into a cohesive argument. Applies excellent critical thinking to evaluate all information sources. Recognizes the utility of different information sources and uses appropriate sources in each context. Understands the counterarguments of a topic and can respond to one's own decision for which is best.</td>
<td>1 Can find sufficient information in various areas. Attempts to evaluate information sources through critical thinking. Gathers data and organizes it in a cohesive argument. May fail to recognize the utility of different sources and/or respond to conflicting ideas.</td>
<td>1 Gathers data but not from the best sources and arguments are lacking the necessary data to make them strong. Seeks information from a limited range of sources. May not evaluate information sources, or may do so without a coherent critical framework.</td>
</tr>
</tbody>
</table>
**Outcome 5:** Exercise a variety of effective communication skills useful for life and ministry.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Communicates information clearly through a variety of media. Message shows consideration of audience, purpose and context. Incorporates affective as well as cognitive elements in communication.</td>
<td>Clearly communicates information through various media. Can use multiple media based on the communication task given.</td>
<td>Communicates basic information through at least some media. Can make some use of at least oral and written communication.</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Written Communication follows standard conventions and is virtually free from grammatical, spelling, and punctuation errors. Uses a variety of sentence structure. Vocabulary reflects mastery in writing and speaking. Incorporates commendable organization of thought evidenced by clearly identifiable introduction, main points, supporting evidence, and conclusion.</td>
<td>Written Communication follows standard conventions and has a few grammatical, spelling, punctuation and sentence structure errors. Vocabulary above elementary level. Includes greater variety in sentence structure. Incorporates clear organization of thought evidenced by apparent introduction, main points, and conclusion. Main points may or may not be sufficiently supported.</td>
<td>Written communication does not meet a satisfactory level of grammatical, spelling, punctuation, and sentence structure. Vocabulary is limited. Lacks variety in sentence structure. Has poor organization of thought with weaknesses in identifying introduction, main points, supporting evidence, and conclusion.</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Demonstrates Oral communication skills that follow standards for good public speaking and creates the desired response.</td>
<td>Demonstrates Oral communication skills that follow standards for good public speaking with a few flaws.</td>
<td>Demonstrates poor oral communication skills.</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>If appropriate, demonstrates electronic communication skills that follow appropriate standards for conventions, aesthetics, variety, consistency, engagement, accessibility, and layering.</td>
<td>If appropriate, demonstrates electronic communication skills that follow appropriate standards for most of the following: conventions, aesthetics, variety, consistency, engagement, accessibility, and layering.</td>
<td>Unable to demonstrate electronic communication skills that follow appropriate standards.</td>
</tr>
</tbody>
</table>
**Outcome 6:** Articulate the value of human beings as created in God's image and support that belief through Christ-like acts of service.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td>Can articulate Scriptural reasons for valuing all human beings. Can address issues of injustice from Scriptural principles. Can identify actions in his/her life that demonstrate putting into practice a belief in all human beings' value. Can explain the importance of creation by a Creator to a Christian theory of human value.</td>
<td>Can explain a belief in the individual value of all human beings. Can identify Scriptural support for this belief. Can discuss how this belief should motivate action in a Christian's life.</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>Able to articulate the importance of service for biblical Christianity. Can identify Scriptural bases for the role of service in the Christian life. Can point to specific, regular practices in his/her life that demonstrate lived commitment to Christian service.</td>
<td>Able to articulate the importance of service for Christians. Can identify Scriptural calls for service in the Christian life. Can point to instances in his/her life that demonstrate a lived commitment to Christian service.</td>
</tr>
</tbody>
</table>
**Outcome 7:** Articulate the essential elements of the Gospel and the Christian faith in a variety of cultural settings.

<table>
<thead>
<tr>
<th>EXEMPLAR</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXEMPLAR**

1. Articulates clearly the essential elements of the Gospel with Scriptural support. Demonstrates the ability to contextualize the Gospel message using different cultural traditions.

2. Will have successfully participated in a range of encounters demonstrating the acquisition of necessary skills and attitudes for cross-cultural communication of the Gospel.

**PROFICIENT**


2. Will have participated in a range of encounters demonstrating the acquisition of basic skills and attitudes for cross-cultural communication of the Gospel.

**DEVELOPING**


2. Will be able to articulate the need for communicating the Gospel cross-culturally.
**Outcome 8:** Employ leadership and other relational skills that reflect a biblical world view in a chosen field of service or vocation.

<table>
<thead>
<tr>
<th>EXEMPLARY</th>
<th>PROFICIENT</th>
<th>DEVELOPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1   Exhibits servant-leadership when given responsibility for leading. Exhibits interpersonal skills such as awareness of cultural differences, perceptiveness, listening, conflict resolution ability, etc. Consistently behaves in accordance with Christian principles of decorum and ethics.</td>
<td>1   Demonstrates ability to lead somewhat effectively but is lacking in some leadership and/or interpersonal skills. Behaves with Christian decorum and ethics.</td>
<td>1   Exhibits potential for leadership, though may lack practical skills. Seeks God's direction for vocation.</td>
</tr>
<tr>
<td>2   Demonstrates an extensive set of knowledge and skills for chosen vocation and conceives of the vocation as ministry. Adapts general principles to specific vocational contexts. Able to articulate a biblical world view of chosen vocation.</td>
<td>2   Demonstrates a core set of knowledge and skills necessary for chosen vocation and views the vocation through a Christian framework.</td>
<td>2   Exhibits a desire to serve God in chosen vocation and exhibits potential for leadership, though may lack practical skills and/or the ability to articulate a biblical world view of chosen profession.</td>
</tr>
</tbody>
</table>
Appendix C: Screenshots of Items for Survey of Stakeholders for Phase Two

Which Bible college are you associated with? (This is the Bible college that supplied your email for this survey. It is listed in your email invitation.)

What is your role in relation to that Bible college?
- Administrator
- Alumnus/Alumna
- Donor
- Faculty member
- Parent of a current student
- Staff member
- Student

What is your gender?
- Male
- Female

What is your age? (Please use numerals.)

What is your ethnicity?
- American Indian or Alaska Native
- Asian
- Black or African American
- Hispanic or Latino

What is the highest level of education you have completed?
- Less than 4-year college degree
- 4-year college degree
- Master’s degree
- Doctorate or professional degree
The following questions present a number of different ways to measure a college's quality. If you want to evaluate how well a college is fulfilling its mission, how important to you is each of these measures?

For each method, please indicate how important the measure is to you for helping you decide whether or not a college is doing a good job of fulfilling its mission.

How important are student scores on a standardized test of Bible-content knowledge?
- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important

How important are student testimonials or biographies?
- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important

How important is a college's completion rate (the percentage of students who start a program at the college and finish it within a reasonable number of years)?
- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important

How important are links to anonymous examples of actual student work?
- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important
How important are scores on student-satisfaction benchmarks from the Noel-Levitz Student Satisfaction Inventory?

- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important

How important are student scores on a standardized test of general education knowledge (e.g., critical thinking, reading, writing, etc.)?

- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important

How important is information about full-time tuition and required fees?

- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important

How important are scores on student-engagement benchmarks from the National Survey of Student Engagement (NSSE)?

- Not important at all
- Not very important
- Somewhat important
- Neither important nor unimportant
- Somewhat important
- Important
- Very important
Appendix D: Screenshots of Sample Items for Survey of Stakeholders for Phase Three
The remaining questions on this survey ask you about sample institutional profiles. The profiles are for imaginary colleges and contain only hypothetical information.

Please use the links below to open all four institutional profiles (these will open in separate Internet browser windows).

After you have opened the profiles, please move to the next page of survey questions.

You may continue to refer to the profiles as you answer questions, so please keep the profiles open in your browser!

Here are links to the four profiles:
- Big City Bible College
- Smallville Bible College
- Suburbiab Institute for Biblical Studies
- Ruralia College
Timing
These page timer metrics will not be displayed to the recipient.

- First Click: 0 seconds
- Last Click: 0 seconds
- Page Submit: 0 seconds
- Click Count: 0 clicks

In case you lose the profiles, here are the links to them again.
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralia College

From the institutional profiles, which institution(s) had a greater percentage of faculty with doctoral degrees than the ABHE average? (Check all that apply.)
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralia College

How confident are you that your answer to the previous question is correct?
- Not confident at all
- Not very confident
- Neither confident or unconfident
- Somewhat confident
- Very confident
Timing
These page timer metrics will not be displayed to the recipient.

- First Click: 0 seconds
- Last Click: 0 seconds
- Page Submit: 0 seconds
- Click Count: 0 clicks

In case you lose the profiles, here are the links to them again.
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralia College

From the institutional profiles, which institution had the largest percentage of students improve their Bible-content test score from their freshman year to their senior year?
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralia College

How confident are you that your answer to the previous question is correct?
- Not confident at all
- Not very confident
- Neither confident or unconfident
- Somewhat confident
- Very confident
Timing
These page timer metrics will not be displayed to the recipient.

- First Click: 0 seconds
- Last Click: 0 seconds
- Page Submit: 0 seconds
- Click Count: 0 clicks

In case you lose the profiles, here are the links to them again.
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralla College

From the institutional profiles, which institution had the lowest percentage of students complete their degrees in the defined time period?
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralla College

How confident are you that your answer to the previous question is correct?
- Not confident at all
- Not very confident
- Neither confident or unconfident
- Somewhat confident
- Very confident
Timing
These page timer metrics will not be displayed to the recipient.
- First Click: 0 seconds
- Last Click: 0 seconds
- Page Submit: 0 seconds
- Click Count: 0 clicks

In case you lose the profiles, here are the links to them again:
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralla College

From the institutional profiles, which institution(s) had the lowest room/board rates?
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralla College

How confident are you that your answer to the previous question is correct?
- Not confident at all
- Not very confident
- Neither confident or unconfident
- Somewhat confident
- Very confident

Percentage slider

Timing
These page timer metrics will not be displayed to the recipient.
- First Click: 0 seconds
- Last Click: 0 seconds
- Page Submit: 0 seconds
- Click Count: 0 clicks

In case you lose the profiles, here are the links to them again.
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralia College

From the institutional profiles, which institution had the best (highest) scores on NSSE (National Survey of Student Engagement) scales?
- Big City Bible College
- Smallville Bible College
- Suburbia Institute for Biblical Studies
- Ruralia College

How confident are you that your answer to the previous question is correct?
- Not confident at all
- Not very confident
- Neither confident or unconfident
- Somewhat confident
- Very confident

Survey Powered By Qualtrics
The questions on this page ask you about specific aspects of the institutional profiles.
For all these questions, satisfaction means that you find the information sufficient for you to judge how well an institution is achieving its mission.

In case you lose the profiles, here are the links to them again:
- Big City Bible College
- Minimally Bible College
- Academic Institute for Biblical Studies
- Burakka College

How satisfied are you with the profile’s information about student scores on a standardized test of Bible-content knowledge?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied

How satisfied are you with the profile’s information about student scores on a standardized test of general-education knowledge?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied

How satisfied are you with the profile’s information about faculty evaluation of seniors’ work as Proactive or Exemplary?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied

How satisfied are you with the profile’s information about faculty degrees/credentials?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied

How satisfied are you with the profile’s links to examples of student work?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied

How satisfied are you with the profile’s student testimonials?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied
<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>How satisfied are you with the profile's alumni testimonials?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
<tr>
<td>How satisfied are you with the profile's completion rate information?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
<tr>
<td>How satisfied are you with the profile's selection rate information?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
<tr>
<td>How satisfied are you with the profile's information about tuition and fees?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
<tr>
<td>How satisfied are you with the profile's information about room and board costs?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
<tr>
<td>How satisfied are you with the profile's information about the student engagement (NSSE scores)?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
<tr>
<td>How satisfied are you with the profile's information about the student satisfaction (SSI scores)?</td>
<td>Not at all satisfied</td>
</tr>
<tr>
<td></td>
<td>Not very satisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Neither satisfied nor dissatisfied</td>
</tr>
<tr>
<td></td>
<td>Somewhat satisfied</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
</tr>
<tr>
<td></td>
<td>Very satisfied</td>
</tr>
</tbody>
</table>
In case you lose the profiles, here are the links to them again.

- Big City Bible College
- Smallville Bible College
- Suburban Institute for Biblical Studies
- Ruralia College

How satisfied are you overall with the profile's information about institutional costs for students?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied

How satisfied are you overall with the profile's information about student learning?
- Not at all satisfied
- Not very satisfied
- Somewhat dissatisfied
- Neither satisfied nor dissatisfied
- Somewhat satisfied
- Satisfied
- Very satisfied