An Online Guide for Evaluating Theological Learning

John Harris
Samford University

This guide is an online complement to the *Handbook of Accreditation*, Section 8, A Guide for Evaluating Theological Learning. It identifies tests and questionnaires that might be used to assess student learning and offers suggestions for selecting assessment artifacts and designing assessment strategies and instruments.

Disclaimer. This review of tests, surveys, and questionnaires is neither comprehensive nor exhaustive. The identified strategies and instruments are only suggested for possible exploration. Neither I nor ATS intends for these suggested strategies and instruments to constitute an approved list. My search for *published* instruments found only a few that a faculty might consider. All suggestions for developing comprehensive tests, surveys, or observation techniques are more about approaches that might be considered; they are not and were not intended to be "blueprints."

Finding Tests and Surveys

There are several very comprehensive and helpful lists with descriptions of tests that could be useful in Master of Divinity (MDiv) student learning assessment.

Measures of Religiosity, Peter C. Hill and Ralph W. Hood Jr., eds., (Religious Education Press, 1999). 5316 Meadow Brook Road, Birmingham, Alabama, 35242. 205-991-1000. Fax: 205-991-9669. Email: releduc@ix.netcom.com.

The chapter headings in *Measures of Religiosity* are as follows:

1. Scales of Religious Belief and Practice
2. Scales of Religious Attitudes
3. Scales of Religious Orientation
4. Scales of Religious Development
5. Scales of Religious Commitment and Involvement
6. Scales of Religious Experience
7. Scales of Religious/Moral Values
8. Multidimensional Scales of Religiousness
9. Scales of Religious Coping and Problem-Solving
10. Scales of Spirituality and Mysticism
11. Scales of the God Concept
12. Scales of Religious Fundamentalism
13. Scales of Views of Death/Afterlife
14. Scales of Divine Intervention/Religious Attribution
15. Scales of Forgiveness
16. Scales of Institutional Religion
17. Scales of Religious Constructs

Measures of Religiosity is valuable in searching for assessments of religious beliefs, attitudes, and behaviors. Reviews of surveys and tests deal with the following topics:

Variable. Describes what the test is intended to assess.

Description. Elaborates on the variable assessed, describes the test's format, gives examples of items and the age range for which it is designed.

Practical Considerations. Deals with administration of the test: who should do what, time required, scoring, and any special issues.

Norms/Standardization. Describes the norm group(s) used to develop the standardized scores.

Reliability. Reports available empirical indicators of the test's consistency or reliability.

Validity. Reports available evidence that the test assesses what it purports to assess, such as its manifest of face validity, correlations with other tests, or how well it predicts certain behavior.

Location. Contact information for obtaining the test.

Recent Research. References of research on the particular test.

References. Related research.

Appendix. A copy of the test or survey reviewed.

After describing the content of the reviews, the editors make a very important disclaimer: "No effort is made to identify the best measures, and reviewers were instructed *not* to provide critical evaluations of the scales; such evaluations are left up to the reader." The editors then suggest six guidelines that should be considered in selecting a test, which are given below:

1. A test should reflect the theoretical concepts that you wish to assess.
2. Check the alignment of the variables reflected in the test items with the variables you want to measure. Put another way, do not use a general religiosity measure to assess or predict particular religious behaviors or vice versa.
3. Consider measures of reliability. According to the editors, "Most measures in this book have at least acceptable internal consistency reliabilities (coefficient alpha), but few are in the excellent range (>.90)."
4. Consider validity indicators. The editors face up to the reality that empirical evidence of validity is very difficult to establish for measures of religiosity. But they believe that content validity can be established by those familiar with the particular religious tradition for which the test is designed.
5. The tests reviewed in this book generally lack the statistical data of more established tests such as the Graduate Record Examinations or the Minnesota Multiphasic Personality Inventory. They often have means and standard deviations; they generally have not been normed against representative samples so that one could say that a given score indicates how an individual score would compare to a national sample.
6. When taking tests of religiosity most of us feel a strong need to respond to items as we think we should rather than how we actually feel.

In addition to these six considerations, most of the tests reviewed do not have standard protocols of administration. That is, for a test to be standardized, it must be introduced and administered to all takers in the same way.

Given all these disclaimers and caveats, should a seminary faculty use *Measures of Religiosity* to identify and evaluate assessment instruments? I believe it should for several reasons. As will be evident in the later sections, I have found few tests even remotely related to ATS's areas of Religious Heritage, Cultural Context, Personal and Spiritual Formation, and Capacity for Ministerial and Public Leadership. To my knowledge, this is the only source for finding assessments of spirituality and religious commitment that may relate to the four goals of the ATS MDiv degree program standard. In most cases, the reviews contain some indicators of reliability that will usually not be available immediately if one develops one's own tests. Similarly, reviewers will share evidence of validity, particularly in terms of observations of experts and sometimes in terms of correlation with other measures.

Measures of Religiosity was originally developed to make those interested in assessing various aspects of religiosity aware of tests faculty and researchers developed but which are generally not nationally published. If a test appears useful, information for contacting the source for the test is provided.

ETS Test Collection (TestLink Overview) contains information on every *published* test and can be accessed without charge on the Internet at: http://www.lib.uaa.alaska.edu/articles/moreinfo.php?item_id=861&contact_id=12.

I used this collection to search for tests and questionnaires that are described later. The ETS collection describes itself as follows:

Abstract. The ETS Test Collection is a library of more than 25,000 tests and other measurement devices that makes information on standardized tests and research instruments available to researchers, graduate students, and teachers. Collected from the early 1900s to the present, ETS Test Collection is the largest such compilation in the world. The tests in this collection were acquired from a variety of U.S. publishers and individual test authors. Foreign tests are also included in the collection, including some from Canada, Great Britain, and Australia. In order to search the ETS Test Collection, you will need Microsoft Internet Explorer version 5.5 or higher or Netscape version 6.0 or higher, with the exception of Netscape version 7.0. Web site: http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=ed462d3631df4010VgnVCM10000022f95190RCRD&vgnextchannel=85af197a484f4010VgnVCM10000022f95190RCRD

The overview descriptions of tests generally come from the ETS Test Collection. The ETS Test Collection TestLink Internet address follows the description. If you wish to review the full ETS Test Collection description, you may do by following these steps:

1. Go to ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes

2. If you are not a subscriber, click on "Public Access."

3. When the search page appears, insert the Title of the test.

When the ETS Test Collection description of a test is cited below, you will be reminded of this process as follows:

ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title

***The Mental Measurements Yearbooks* (MMY)** are indispensable in reviewing tests. They have been published every eighteen to twenty-four months since 1938; the latest was released in 2005. Each test listed in the MMY has two independent reviews. These reviews include empirical evidence of validity and consistency or reliability. While the reviews include technical terms and statistics, they generally conclude with observations useful for nonspecialists. A search for MMY reviews was made for each test found in the ETS Test Collection; when MMY reviews were found, they are briefly noted.

The Buros Institute of Mental Measurements describes the MMY as follows:

> Typical MMY test entries include descriptive information, one or two professional reviews, and reviewer references. To be reviewed in the MMY, a test must be commercially available; be published in the English language; and be new, revised, or widely used since it last appeared in the MMY series. Beginning in The Fourteenth Mental Measurements Yearbook, tests also must provide sufficient documentation supporting their technical quality to meet criteria for review. Between the years 1988 and 1999, the Buros Institute produced Supplements to the Ninth through Thirteenth Mental Measurements Yearbooks. Beginning with the Fourteenth Mental Measurements Yearbook, volumes in the MMY series will be produced every 18 to 24 months and the Supplements will be discontinued.[1]

***Electronic Access to Mental Measurements Yearbook Series*.** Most university libraries have online subscriptions for The Mental Measurements Database. One can also access MMY through the Buros Center for the Testing Web site: www.unl.edu/buros.

> "Test Reviews Online" is a Web-based service of the Buros Institute of Mental Measurements. Test reviews are available to individual users exactly as they appear in the Ninth through Sixteenth Mental Measurements Yearbook series. In addition, monthly updates are provided from our latest test review database. For a small fee, users may download reviews for more than 2,500 tests that include specifics on test purpose, population, publication date, administration time, and descriptive test evaluations. The Mental Measurements Yearbook Database currently is available from EBSCO and OVID/SilverPlatter at many academic libraries. Containing the text of the Ninth through the Sixteenth Mental Measurements Yearbooks, the MMYD is updated every six months to ensure timely access to current information. The MMYD is available to libraries and organizations by subscription and may be provided as a stand alone service for authorized users.[2]

Tests in Print, published by The Buros Institute of Mental Measurements, is a valuable search tool for tests and surveys; it is described as follows:

> *Tests in Print* (TIP) serves as a comprehensive bibliography to all known commercially available tests that are currently in print in the English language. Now in its sixth edition, TIP provides vital information to users including test purpose, test publisher,

> in-print status, price, test acronym, intended test population, administration times, publication date(s), and test author(s). A score index permits users to identify what is being measured by each test. Tests in Print also guides readers to critical, candid test reviews published in the Mental Measurements Yearbook series. *Tests in Print* is an indispensable reference for professionals (including such areas as education, psychology, business) and anyone interested in the critical issues of tests and testing.[3]

***ATS Online Assessment Resources*.** James Meek developed an extensive list of online assessment sites with descriptions of what each offers and with links to each of them. Meek summarizes his list as follows:

> There are many helpful assessment resources available on the Web, prepared by all kinds of institutions and organizations. The principles and practices of assessment are the same, whether one is training artists, engineers, or theologians. The Web sites have been selected as those most likely to be useful to those involved in theological education.

One can access the list of resources directly: http://www.ats.edu/projects/OnlineAssessmentResources-Basics.asp.

RELIGIOUS HERITAGE goal of MDiv Degree Program Standard A

The ATS Commission on Accrediting's MDiv Degree Program Standard A, Section 3.1.1 requires that students demonstrate a comprehensive and discriminating understanding of the religious heritage of their faith community. In particular, they should demonstrate understanding of

- Scripture
- historical development and contemporary articulation of the doctrinal, theological tradition of the community of faith
- social and institutional history of that community
- the broader heritage of the Christian tradition
- specific character of particular Christian traditions and communities
- ways in which the traditions transcend particular social and cultural settings
- interdependence of all the above areas of Religious Heritage with Cultural Context, Personal and Spiritual Formation, and Capacity for Ministerial and Public Leadership.

Possible Instruments

DANTES Subject Standardized Tests: Introduction to World Religions. From my search of the ETS Test Collection, only the DANTES Subject Standardized Tests: Introduction to World Religions appears to offer any value in assessing any aspect of Religious Heritage as described in the ATS Standards. This test is not reviewed in *Mental Measurements Yearbook*. The ETS Test Collection describes it as follows:

Abstract. The DANTES program is a series of secure tests administered by postsecondary institutions to grant credit by examination for education gained outside the classroom. Examinations may be worth from two to six credit hours in a baccalaureate program, baccalaureate upper division program, or a technical program. A minimum score for credit has been established by the American Council on Education. Individual institutions administer the examinations as well as set the fees and schedules. These instruments complement the College Board's College Level Examination Program (CLEP) with several instruments in applied technology. They were originally developed for military personnel. Major areas of assessment include mathematics, social science, physical science, business, foreign language, and applied technology. For each test, a fact sheet containing the curriculum specifications of the course, a list of texts on which the test is based, and statistical information is available.

Tests are revised regularly, but generally the following topics commonly taught in courses on this subject are covered in this examination: dimensions and approaches to religion, primal religions, Hinduism, Buddhism, Confucianism, Taoism, Judaism, Christianity, Islam.

ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title

Availability. Educational Testing Service, DANTES Program Office, Princeton, NJ 08541.

Comment. This DANTES Introduction to World Religions test is a general education test for undergraduates. A seminary *might* find it helpful in assessing entering students' basic knowledge of world religions.

Episcopal Church General Ordination Examination. Professional schools, such as law and pharmacy, often use licensing examination results as one way of assessing outcomes. Some denominations, such as the Episcopal Church and the Presbyterian Church (U.S.A.), have extensive ordination examinations that could be used to assess seminary learning outcomes. The Episcopal Church's General Ordination Examination (GOE) covers the following Seven Canonical Areas:

I. The Holy Scriptures
II. Church History, including the Ecumenical Movement
III. Christian Theology, including Missionary Theology and Missiology
IV. Christian Ethics and Moral Theology
V. Studies in Contemporary Society, including Racial and Minority Groups
VI. Liturgics and Church Music
VII. Theory and Practice of Ministry[4]

To write the GOE requires four full days. For example, the 2005 GOE followed the schedule as shown below.

General Ordination Exam Schedule for 2005

Day	**Time**	**Examination Area**
Monday	9 a.m.–12:30 p.m.	Contemporary Society
Monday	1:30–5 p.m.	Liturgy and Church Music
Tuesday	9 a.m.–12:30 p.m.	Holy Scripture
Tuesday	1:30–5 p.m.	Church History

Thursday	9 a.m.–5 p.m.	Christian Theology
Friday	9 a.m.–12:30 p.m.	Christian Ethics and Moral Theology
Friday	1:30–5 p.m.	Theory and Practice of Ministry

To review the 2005 GOE see: http://www.episcopalchurch.org/1521_60754_ENG_Print.html.

Comment. These are well-constructed essay examinations that integrate knowledge, understanding, and application. A profile of graduates' GOE performance could be developed each year and used as one indicator of a seminary's learning outcomes. Furthermore, the general approach and format of these essay examinations could be very helpful to faculty committees charged with writing comprehensive examinations of Religious Heritage learning.

The Presbyterian Church (U.S.A.) Ordination Examinations. The PC(USA) standard ordination examinations deal with the following five areas: (1) Open Book Bible Exegesis, (2) Bible Content, (3) Theological Competence, (4) Worship and Sacraments, and (5) Church Polity.[5] "The purpose of ordination examinations is to determine one's entry-level readiness for ministry in the Presbyterian Church (U.S.A.). The exams present pastoral challenges whereby candidates can demonstrate how they integrate faith and theological education in practical ministry."[6] The PC(USA) Presbyteries' Cooperative Committee on Examinations for Candidates Web site provides a complete description of the Ordination Examinations, http://www.pcusa.org/exams/index.htm. Copies of previous years' examination are also available on request.

Comment. In reviewing the February 2005 examinations in Bible Exegesis, Theological Competence, Worship and Sacraments, and Church Polity, I was struck by how well they integrate knowledge, understanding, and application. It seems to me that PC(USA) seminaries would find them extremely helpful in assessing learning outcomes. They could be treated as artifacts for assessing some MDiv outcomes.

The Presbyterian Church (U.S.A.) Bible Content Examination. The Presbyteries' Cooperative Committee on Examinations of the PC(USA) also provides the Bible Content examination.

> The Bible Content examination is intended to assess one's knowledge of stories, themes, and pertinent passages in the Old and New Testaments. This exam is ordinarily taken in the first year of seminary (*Book of Order*, G-14.0310b). The two-hour examination contains 100 multiple-choice questions about the contents of the Bible, using the New Revised Standard Version for all quotations. The exam is given only once a year. The questions vary in difficulty, but cover the entire Bible as required by action of the General Assembly in 1978.[7]

Complete copies of each year's Bible Content Examinations are available online at this Web site: http://www.whitneyhq.com/biblecontent/. This Web site allows anyone to take the examination and receive a score of the number answered correctly.

Comment. The PC(USA) Bible Content Examination is a test of factual knowledge of the Bible with 100 questions distributed over the following division of the Bible: The Pentateuch, Historical Books, Prophets, Psalms and Wisdom Literature, Gospels, Acts and Pauline Letters, and Rest of New Testament (1 John–Revelation). Having this examination online with online scoring is a valuable resource for seminaries interested in checking their students' Bible knowledge.

Standardized Bible Content Test. The Standardized Bible Content Test was developed and published by: Association for Biblical Higher Education, 5575 S. Semoran Blvd., Suite 26, Orlando, FL 32822-1781. 407-207-0808. Email: info@abhe.org.
It is not listed in the ETS Test Collection nor reviewed in Mental Measurements Yearbook.

The Standardized Bible Content Test (SBCT) was developed to assess Bible college students' knowledge of the Bible. It is a knowledge test in the sense of asking students to identify correct facts from the Bible. All questions are five-option, multiple choice items. There are several forms or editions of the SBCT varying in number of questions from fifty to 150. Form G has 150 items with seventy-five from the Old Testament and seventy-five from the New Testament. In each testament, the items are distributed as follows:

Introduction	7
People	23
Books	23
Quotes	11
Places	11
Total	75

Comment. The SBCT offers a quick way to check on students' basic recall of Bible facts. The SBCT Manual demonstrates the SBCT was developed and has been refined for its intended coverage. It has been refined by appropriate psychometric and statistical tools. The manual provides extensive item analyses. As with any test, a faculty would need to determine if the SBCT assesses the knowledge expected of students. My review of the Manual suggests that the reliability of the test has been empirically researched and demonstrated. If seminaries develop their own comprehensive objective tests to survey their students' knowledge, they would do well to have someone analyze the test and its constituent items as the SBCT test designers have done.

Locally Developed Comprehensive Tests

Some will know of tests of biblical and theological knowledge and understanding that would interest others in the ATS community. When this is the case, please let any member of the accrediting staff know of them. Brief descriptions and contact information will be posted on the ATS Online Assessment Resources Web page.

If a faculty chooses to develop its own comprehensive test for Religious Heritage, it might adopt a test plan for developing items in proportion to the goals. Items could be developed in certain content areas such as

1. Scripture
2. historical development and contemporary articulation of the doctrinal theological tradition of the community of faith
3. social and institutional history of that community
4. broader heritage of the Christian tradition
5. specific character of particular Christian traditions and communities
6. ways the traditions transcend particular social and cultural settings

7. interdependence of all the above areas of Religious Heritage and with Cultural Context, Personal and Spiritual Formation, and Capacity for Ministerial and Public Leadership

In addition to content categories, a second dimension of knowledge or understanding could be added. Shulman's "six levels of learning" might be considered.[8] On the other hand, a faculty might write test items for only three levels of learning: Knowledge (recall of facts), Comprehension (capable of stating a concept in one's own words), and Understanding (ability to apply factual and conceptual knowledge to a variety of circumstances).

A test-item plan could be developed as a matrix with seven columns for the seven ATS Religious Heritage areas, and three rows for three levels of performance.

Religious Heritage Test-Item Development Plan

Religious Heritage Areas / Levels	1 Scripture	2 Doctrine & Theology	3 Social & Inst. History of Faith Comm.	4 Broader Christian Tradition	5 Specific Character Christian Traditions & Comm's.	6 How Traditions Transcend Societies & Cultures	7 Integration 1-6
A Knowledge	**1A** Items	**2A** Items	**3A** Items	**4A** Items	**5A** Items	**6A** Items	**7A** Items
B Comprehension	**1B** Items	**2B** Items	**3B** Items	**4B** Items	**5B** Items	**6B** Items	**7B** Items
C Understanding	**1C** Items	**2C** Items	**3C** Items	**4C** Items	**5C** Items	**6C** Items	**7C** Items

The seven column headings in the table above are derived from the ATS MDiv Degree Program Standard A, Section 3.1.1 Religious Heritage, A.3.1.1.1., A.3.1.1.2., and A.3.1.1.3 as indicated below:

A.3.1.1 *Religious Heritage*: The program shall provide structured opportunity to develop a ***comprehensive and discriminating understanding*** of the religious heritage. (Emphasis added.)

A.3.1.1.1 Instruction shall be provided in Scripture **(1. Scripture)**, in the historical development and contemporary articulation of the doctrinal and theological tradition of the community of faith **(2. Doctrine & Theology)**, and in the social and institutional history of that community **(3. Social & Inst. History of Faith Comm.)**.

A.3.1.1.2. Attention should be given both to the broader heritage of the Christian tradition as such **(4. Broader Christian Tradition)** and to the more specific character of particular Christian traditions and communities **(5. Specific Character Christian Traditions &**

Comm's.), to the ways the traditions transcend particular social and cultural settings, and to the ways they come to unique expression in them **(6. How Traditions Transcend Societies & Cultures)**.

A.3.1.1.3. Instruction in these areas shall be conducted so as to indicate their interdependence with each other and with other areas of the curriculum, and their significance for the exercise of pastoral leadership. **(7. Integration 1-6)**.

Of the three cognitive levels suggested in the Test Plan, "understanding" might be emphasized given the aim of *"comprehensive and discriminating understanding.*"

This plan calls for writing questions for each of the cells. Such a plan helps a faculty to determine the percentage of the test that it will devote to each of the seven components of Religious Heritage at three performance levels. A faculty can obviously decide the percentage of the test to be devoted to each content area (each column) and the percentage of items for each level of performance (each row). The percentage for each is determined by multiplying its column and row percents.

Faculty members could write questions for their respective content areas. To encourage consistency in writing items, consider providing examples of items for assessing each of the performance levels.

Before giving such a test to students, it might be reviewed in two ways:

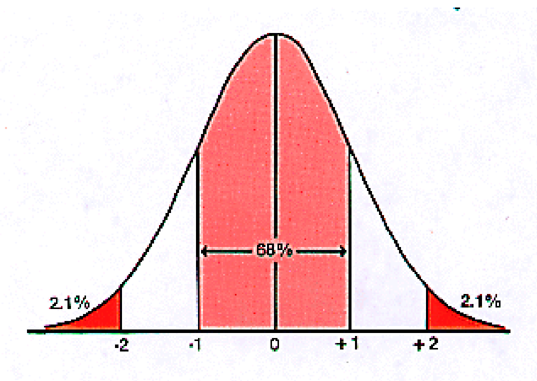
1. Two or three theological educators from another seminary might review the items against their respective goals. For this review, the items might be clustered by cells; for example, all items in cell 2B deal with "the historical development and contemporary articulation of the doctrinal and theological tradition of the community of faith" at the Understanding level.

2. Also, consider pilot-testing the test by asking three or more individuals to actually respond to it.

Test-item banks will prove very useful if the test is to be given to each graduating class. Over time, multiple items can be written and filed for each item. For such a bank, items can be filed by cell (e.g., 1A, 1B, etc.). Such efforts deserve to be sustained and refined over several years. The quality of each item could be assessed each time it is used to improve or eliminate it. And new items could be added each year. Eventually, the faculty will have a pool of items for each cell.

<u>Evaluating Test Items</u>. There is a substantial body of theory and research on evaluating tests and their component items; a comprehensive review of it is beyond the scope of this document. Nevertheless, there are some principles to offer.

In many cases, nationally standardized achievement tests and in some cases, locally developed achievement tests, are built on the assumptions of ability tests. Ability tests are designed to determine differences among individuals. Ideally, they are constructed to distribute a significant number people over a normal distribution. Almost any variable in nature, if extensively measured, will produce a normal curve. Human ability tests, such as IQ tests and the Miller's Analogy Test, are designed to spread individuals across a normal curve of distribution with about two thirds of the individuals hovering around the average score and extending to individuals at the extremes of the range, below and above the average. The following graphic depicts a normal distribution:



Items for ability tests are evaluated in terms of how well they discriminate. The ideal item in such test is one that 50 percent of those taking the test answer correctly and, consequently, 50 percent fail. Obviously, as a test designer continues to eliminate items that do not discriminate at this 50 percent level, the resulting test emphasizes differences among individuals so that a normal curve results.

This approach is not appropriate when the purpose of a test is to compare the effect of a course or instructional program on a group of students or to determine the number of students who reach a predetermined level of mastery. Obviously, if one keeps adding items that only 50 percent of the respondents get right, you are measuring how they differ in native intelligence—not the basic understandings, knowledge, and skills they have in common as the result of a given course or instructional program. Such tests are more sensitive to individual abilities than instructional effects. This is a major reason assessment reveals no significant difference between one curriculum and another or between one approach to teaching and another. The tests are sensitive to individual differences not group differences.

To assess the effectiveness of one curriculum or instructional program, "treatment-sensitive" tests are needed. Rather than judging and selecting items by how well they discriminate among individual differences, we need items validly derived from the learning goals of the course. Such items should then be evaluated by how well they discriminate between those who have mastered particular concepts or skills and equally intelligent individuals who have not. There are several ways to gauge items for inclusion in tests designed to assess predetermined levels of mastery to be sensitive to particular curricula or instructional methodologies.

To determine how well an item reflects a desired learning outcome, a jury of competent individuals can compare each item to its goal and rate how well each item reflects its goal. Such a jury could also take the test, and individuals of similar general abilities but not proficient in the test's particular content could also take the test. This allows one to determine the items that discriminate among informed and uninformed individuals of comparable general ability.

To see if items are sensitive to a particular instruction program, a test could be given to students on the first day and given again at the end of the course, perhaps as the final examination. This allows one to identify the items that distinguish before and after instruction. The items that few, if any, students answered correctly on the first day of class that most, if not all, answered correctly on the final examination reflect the results of the instruction. That is, these are the items that are sensitive to the instruction. The items that do not discriminate between before and after instruction are not sensitive to the instruction.

Items in tests designed to maximize differences among individuals that all the takers answer correctly or incorrectly are eliminated from the test. These items do not add value in maximizing ability differences. If the purpose of a test is to determine mastery of a goal, eliminating items that validly assess the learning goal but do not discriminate among individuals is unjustified. In tests to determine whether students as a group have reached certain goals or to assess differences in curricular and instructional outcome, each item's value is determined by how well it discriminates between experts in a given area compared to individuals of similar abilities and between the instructed and the uninstructed.

Atomistic and Holistic Marking. When grading papers, we typically use atomistic marking. That is, we note specific errors of logic, fact, and grammar. Essentially, we edit papers to suggest improvements as well as determine grades. Holistic grading is another way to evaluate papers. Holistic assessment assumes according to Lindemann "that all the features of a composition or all the skills that comprise writing ability are related, interdependent."[9] Lindemann also cites Charles Cooper's description of the holistic approach.

> Holistic evaluation of writing is a guide procedure for sorting or ranking written pieces, The rater takes a piece of writing and either (1) matches it with another piece in a graded series of pieces or (2) scores it for the prominence of certain features important to that kind of writing or (3) assigns it a letter or number grade. The placing, scoring, or grading occurs quickly, impressionistically, after the rater has practiced the procedure with other raters. The rater does not make corrections or revisions in the paper. Holistic evaluation is usually guided by a holistic scoring guide that describes each feature and identifies high, middle, and low quality levels for each feature.[10]

Holistic evaluation can be a valid, reliable, and time-saving way for a team of readers to assess artifacts. Reliability or consistency among readers is increased when they agree on criteria for ranking or sorting papers. Using these criteria, a group or jury practices marking papers and then comparing similarities and differences in their respective rankings. Those who teach rhetoric or composition often use holistic assessment and know the process for norming or calibrating. Citing Charles R. Cooper and Lee Odell, Lindemann offers a concise and useful summary of the norming process.[11] Given such training, Lindemann describes holistic ranking's efficiency and reliability as follows:

> Because raters don't stop to mark the paper, to correct or revise the student's work, they can score a large number of papers in a short time, "spending *no more than two minutes* on each paper." And because they have practiced matching each against predetermined criteria, they "can achieve a *scoring reliability as high as .90 of individual writers*."[12]

The "two minutes" to mark a paper may be sufficient for the type of essays done in undergraduate composition classes and may not apply to longer seminary papers. Nevertheless, holistic ranking generally requires much less time than atomistic marking. The .90 reliability coefficient may not be typical, but holistic ranking preceded by norming practice results can be quite consistent.

Religious Heritage Assessment Summary. My search for nationally standardized and reviewed examinations for Religious Heritage essentially identified none. Ordination examinations of the caliber of those of the Episcopal Church and PC(USA) could have a key role in assessing MDiv outcomes. Given the lack of nationally recognized examinations and only a few comprehensive ordination examinations, seminaries or seminary consortia might develop their own Religious Heritage comprehensive examinations. There are many ways to plan a test to reflect systematically learning goals; an elementary test design plan was presented. There is an important difference in how ability and outcomes tests are built. Items in outcomes test should be selected according to how they discriminate (1) between those who know and do not know, (2) between those who have been instructed and have learned and those who have not been instructed and have not learned, and (3) between generally intelligent individuals and those with discernable mastery of a subject. Ability tests are designed to discriminate among individuals to yield normal distributions. Finally, efficiency and reliability favor holistic rather than atomistic marking of artifacts to estimate how well a group of students reach MDiv learning goals.

CULTURAL CONTEXT goal of MDiv Degree Program Standard A

In searching the ETS Test Collection for Cultural Context tests, only the following three seem to be somewhat related to cultural context.

Central West Virginia Cultural-Awareness Quiz. The ETS Test Collection abstract describes this test as follows:

Abstract. Designed to help supervisors, managers, and others who may use test results to experience the frustration and confusion that a culturally different person may experience during a testing situation. Said to help others understand the lifestyles of those from other cultures. Items consist of colloquial expressions or local names for common items. Respondent selects the meaning from a choice of four or selects the standard English equivalent. Part of a large kit of instruments that may be reproduced as needed (up to 100 copies) without permission. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. University Associates, 8517 Production Avenue, San Diego, CA 92121.

Comment. Buros Mental Measurement Year Books (MMY) has no review of Central West Virginia Cultural-Awareness Quiz.

Culture Shock Inventory. The ETS Test Collection abstract describes this test as follows: “Designed to assess cross-cultural attitudes. May be used with those who expect to work outside their own culture or work with people from other cultures.” *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Organizational Tests (Canada) Ltd.; PO Box 324; Fredericton, NB, Canada E3B 4Y9

Comment. Not reviewed in MMY.

Intercultural Values Inventory. The ETS Test Collection abstract describes this test as follows:

Abstract. This revised version of the original instrument by Robert L. Kohls (1981) is a forced-choice scale designed to assess cultural values that characterize various racial or cultural groups. It is based on the value orientations proposed by Kluckhohn and Strodtbeck (1961). May be administered to anyone who has completed a ninth-grade education. Norms are available for Black and White subjects. Technical data on reliability for each subscale and validity are available. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Tests in Microfiche: Test Collection, Educational Testing Service, Princeton, NJ 08541.

Comment. After reviewing a copy of the *Intercultural Values Inventory* secured through the microfiche source cited above, some seminaries might find it useful. It assesses five value orientations: human nature, person/nature, time, activity, and social relations. There are thirty questions for each of the five orientations for a total of 150 items. Through the source above, one can obtain a review of the test by Robert T. Carter at Teachers College, Columbia University, and Janet E. Helms, University of Maryland.

While the three instruments cited above might be useful to seminaries, they are not directly related to my understanding of the Cultural Context goal stated in the MDiv Degree Program Standard. Cultural awareness in the standard appears to deal with developing in MDiv students the ability to read or exegete a congregation's or ministry's own culture and the culture of its constituents and contextual community.

Two scholars, a sociologist of religion and a cultural anthropologist, both involved in educating ministers, were asked for guidance and perspective in assessing cultural context. Penny Long Marler,[13] known for her work in congregational studies, offered the following suggestions:

> . . . culture is notoriously difficult to measure. At least in my discipline, the most reliable approaches are ethnographic and participant observation, focus groups, and ethnographic interviews. Triangulation is the best methodology (ethnographic interviewing, participant-observation, and surveys).[14]

Marler recommends Carl Dudley and Nancy Ammerman's summary of tools to assess congregational culture,[15] especially Ammerman's chapter on Culture/Ethnography. Her syllabus, with its references and assignments that integrate theory and on-site involvement, appears to be very helpful.

Sherwood Lingenfelter,[16] a cultural anthropologist and provost at Fuller Theological Seminary, offered his understanding of the ATS MDiv Cultural Context goal and suggested how it might be assessed.

> A.3.1.2 *Cultural Context:* The program shall provide opportunity to develop an understanding of the cultural realities and structures within which the church lives and carries out its mission.
>
> Possible Learning Outcomes:
> 1. Graduates have observed and reflected upon the tension between the need of local congregations to be both contextual and missional.

2. Graduates have compared how two or more local congregations define their place in the wider cultural community and engage in missional outreach.

Assessment Options:
1. Where in the curriculum have these observations and reflections occurred? How have faculty and students (survey) evaluated student learning from these learning projects?
2. Graduating student comprehensive exam: how do you understand God's mission for the church in the twenty-first century?

A.3.1.2.1 The program shall provide for instruction in contemporary cultural and social issues and their significance for ministry. Such instruction should draw on the insights of the arts and humanities, the natural sciences, and the social sciences.

Possible Learning Outcomes:
1. Graduates have observed and reflected upon how congregations engage or ignore one or two critical socio-economic issues of their community.
2. Graduates envision congregational ministries that address the spiritual and socio-economic issues of their local communities.

Assessment Options:
1. Small groups of graduating students work together to present to faculty their assessment of the most critical issues facing local congregations and reflections on how to address these issues. Faculty members assess group work with reference to learning outcomes sought.
2. Student portfolios could include individual or group papers from courses that have addressed the learning outcomes defined above.

A.3.1.2.2 MDiv education shall address the global character of the church as well as the multicultural and cross-cultural nature of ministry in North American society and in other contemporary settings. Attention should also be given to the wide diversity of religious traditions present in the social context.

Possible Learning Outcomes:
1. Graduates have read ten key works on the global character of the church and have reflected upon the implications of these trends for local congregational ministry.
2. Graduates have observed, participated in, or served as an intern in a congregation of different cultural and theological persuasion than their own and have reflected upon the meaning of diversity for world Christianity.

Assessment Options:
1. Requirements for student reading in the curriculum and reflection assignments.
2. Requirements for student internship or practicum and reflection papers reviewed by faculty or peers for quality of learning.[17]

Sensitivity to prejudicial thinking and behaving could be observed by seminary "listening posts." Persons serving as observers or listeners do not report on individuals or isolated events. Rather they are asked to report on patterns of behavior that they observe over time. Listening posts might be selected to represent a diagonal slice of the organization. A diagonal representation means that all layers and divisions of the seminary are represented. Cooks, custodians, secretaries, and security staff often see behaviors that are less visible to faculty and administrators. In addition to strategic observers, two other types of observers might be considered. First, visitors, new employees, and new students will often see things that regulars may not notice. Second, cultural anthropologists might observe a seminary community and report its pervasive attitudes and values.

If students' abilities to read a culture are to be observed in their field experiences, consider the following suggestions:

1. Involve two or more observer-reporters.
2. Identify the types of behaviors with examples that they should look for.
3. Stress the importance of common patterns rather than unusual events.
4. Provide practice, feedback, and discussion among the observers to encourage more inter-consistency.

If discerning cultural context is to go beyond a general sharpening of one's awareness and reflective, tacit understanding, MDiv students might be introduced to the quantitative and qualitative methods commonly used in nearly every aspect of civic, business, and organizational life. This introduction could include basic understanding and skill in the following areas:

1. Survey/questionnaire design, analysis, and interpretation
2. Elementary, descriptive statistics (e.g., mean, median, mode, distribution curves, standard deviation, percentile scores)
3. Sampling
4. Ethnographic techniques
5. Focus groups

<u>Cultural Context Assessment Summary</u>. The search for questionnaires to assess seminarians' abilities to read the culture of a congregation or its communal context turned up only three instruments. Based only on brief summaries of them, they do not appear to address directly the learning about cultural context as indicated by the Commission standard. Marler's work provides sound guidance, and Lingenfelter offers a useful understanding of cultural context and specific ways it might be addressed and assessed. Perhaps, Lingenfelter, Marler, and similar scholars could propose and develop specific strategies and instruments for observing and assessing seminarians' ability to read congregational and community cultures.

PERSONAL AND SPIRITUAL FORMATION goal of MDiv Degree Program Standard A

The content area for Personal and Spiritual Formation reads as follows in the standard:

Personal and Spiritual Formation A.3.1.3
The program shall provide opportunities through which the student may grow in personal faith, emotional maturity, moral integrity, and public witness. Ministerial preparation includes concern with the development of capacities—intellectual and affective, individual and corporate, ecclesial and public—that are requisite to a life of pastoral leadership.

> A.3.1.3.1 The program shall provide for spiritual, academic, and vocational counseling, and careful reflection on the role of the minister as leader, guide, and servant of the faith community.
>
> A.3.1.3.2. The program shall provide opportunities to assist students in developing commitment to Christian faith and life (e.g., expressions of justice, leadership

> development, the devotional life, evangelistic witness) in ways consistent with the overall goal and purpose of the school's MDiv program.

Given my research for this section, a search for assessments for Personal and Spiritual Formation might well begin and end with *Measures of Religiosity* described at the beginning of this article.

The study of the relationship of health to religious and spiritual life has led researchers to look for and, in several cases, develop reliable and valid measures of religiosity and spirituality. Two comprehensive reviews of such measures are described below.

"Scientific Measures of Religiosity and Spirituality: Assessing a Missing Component in Health Communication Research" by Nichole Egbert, Jacqueline Mickley, and Harriet Coeling is available at: http://www.leaonline.com/doi/abs/10.1207/S15327027HC1601_2;jsessionid=iJWNN8cU08X9e3FjfH?cookieSet=1&journalCode=hc

> This article reviews a sample of the major empirical instruments used in this research ("the relation between religious or spiritual variables and health outcomes"), including extrinsic and intrinsic religiosity, spiritual well-being, and religious coping. …[18]

Multidimensional Measurement of Religiousness/Spirituality for Use in Health Research: A Report of the Fetzer Institute/National Institute on Aging Working Group[19] reviews assessments and provides extensive bibliographies in the following areas of religious and spiritual beliefs, values, and practices:

- Daily Spiritual Experiences
- Meaning
- Values
- Beliefs
- Forgiveness
- Private Religious Practices
- Religious/Spiritual Coping
- Religious Support
- Religious/Spiritual History
- Commitment
- Organizational Religiousness
- Religious Preference

The report is available at: http://www.fetzer.org/PDF/Total_Fetzer_Book.pdf.

A search of the ETS Test Collection identified several surveys and tests that might be useful based on their abstracts. The Test Collection abstracts are copied below, and complete Test Collection entries on these instruments, including the copied Abstracts and Availability (sources) may be found by going to the following Web site and inserting the name of the survey or test: *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

When there is a *Mental Measurement Yearbook* entry of a survey or test reviewed in quotes from and comments on the MMY, reviews are included.

Measure of Materialistic Attitudes

Abstract. The Measure of Materialistic Attitudes (MMA) scale defines materialistic attitudes as orientations emphasizing possessions and money for personal happiness and social progress. The MMA is composed of six Likert-type items scored on a five-point disagree-agree scale. Item scores are summed to form an overall MMA index. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Bearden, William O., et. al., *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research* (Newbury Park, CA: Sage Publications, 1993).

Comment. Not reviewed in MMY.

Richins' Materialism Measure

Abstract. Richins' Materialism Measure defines materialism in terms of its role in consumer culture as . . . "the idea that goods are a means to happiness; that satisfaction in life is not achieved by religious contemplation or social interaction, or a simple life, but by possession and interaction with goods." The scale consists of six items measured on a seven-point Likert scale from strongly disagree to strongly agree. It measures two factors: personal materialism and general materialism. Item scores are summed within factors to form indices for each factor. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Bearden, William O., et. al., *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research* (Newbury Park, CA: Sage Publications, 1993).

Comment. Not reviewed in MMY.

Spiritual Checkup

Abstract. The Spiritual Checkup was designed as a checklist to assess spiritual self-care habits. It consists of ten questions rated on a five-point, Likert-type scale. It looks at participation in a religious life, meditation, prayer, reading inspirational materials, and inspiration through art, music, and dance. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Whole Person Associates, 210 W. Michigan, Duluth, MN 55802-1908. 800-247-6789. Web site: www.wholeperson.com.

Spiritual Transcendence Scale

Abstract. The Spiritual Transcendence Scale (STS) was designed to measure spiritual qualities of individuals. It has three subscales: universality (a belief in the unity and purpose of life); prayer fulfillment (an experienced feeling of joy and contentment that results from prayer and/or mediation) and Connectedness (a sense of personal responsibility and connection to others). It consists of twenty-four items rated on a five-point, Likert-type scale. A rater version is to be completed by peer evaluators and is worded in the third person. It has been used with college students in the United States and in India with

Christians, Muslims, Jews, and Hindi. Reliability and validity are discussed. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. *Journal of Personality*, 0022-3506, 67 no. 6, 985-1013, Dec 1999 and article reprint. *American Behavioral Scientist*, 0002-7642, 45 no. 12, 1888-901, Aug 2002.

Comment. Not reviewed in MMY.

Spiritual Transformation Inventory

Abstract. The Spiritual Transformation Inventory (STI) is a measure of Christian spirituality developed by Todd W. Hall, PhD. The purpose of the STI is to help you better understand your spiritual strengths and growing edges at this point in your spiritual journey in order to facilitate intentionally pursuing the next step in your spiritual transformation process. The initial norms are based on a sample of 372 Christian college students. The individual scales and domains were developed through psychometric analyses to verify that items on a scale measure a unitary concept. Individual scales were further clustered into domains using psychometric analyses. Associations with numerous related variables were examined, corroborating that the scales measure what they are intended to measure. Internal consistency was adequate across all the scales. . . . The STI is based on a *relational* and *multidimensional* model of spirituality. The overarching concept permeating the STI is that we are transformed *through* loving relationships with God and others and *for* loving relationships with God and others. Within this broad framework, there are multiple destinations or end goals of spiritual maturity, and multiple pathways through which transformation occurs.[20]

The STI Scale Summary includes the following:

1. Spiritual Meaning and Vitality
   Prayer Type Frequency Scale
   Spiritual Practices Scale-Frequency
   Desolation/Consolation with God Scale
   Transformational Suffering Scale
   Spiritual Perspective Scale
   Spiritual Meaning Scale
   Spiritual Openness Scale
   Awareness of God Scale
   Realistic Acceptance Scale
   Other-Centered Love Scale
2. Spiritual Commitment and Community
   Spiritual Service Scale
   Faith Centrality Scale
   Spiritual Community Scale
3. Secure Spiritual Attachment
   Secure God Attachment Scale
   Forgiveness with God and Others Scale
4. Preoccupied Spiritual Attachment
   Preoccupied God Attachment Scale
   Disappointment Scale
   Instability Scale
5. Dismissing Spiritual Attachment

Compensatory Spiritual Confidence
Spiritual Practices Scale-Motives
Quest as Ultimate Goal Scale

6. Additional Scales and Open-Ended Questions
   Spiritual Hindrances Questions
   Recent Crisis Questions
   Curricular & Co-Curricular Impact Scale

Availability. Concentus Assessment Solutions, Inc., PO Box 5464, Whittier, CA 90607. 562-695-9300 or toll free 877-695-9300. Fax: 562-695-9199. info@concentusassessment.com

Comment. STI is not listed in the ETS Test Collection nor reviewed in Mental Measurement Yearbook. It was developed by Todd Hall, a faculty member at Rosmead School of Psychology, Biola University, who holds a PhD in clinical psychology and a master's in measurement and psychometrics. To date, STI comparison statistics are derived from its use by member institutions of the Council for Christian Colleges and Universities. There are no seminary norms. STI can be taken and scored online.

In addition to the STI, Hall has published two papers that seminary faculty might find useful in planning assessment of MDiv outcomes:

> Todd Hall, "The Personal Functioning of Pastors: A Review of Empirical Research with Implications for the Care of Pastors," *Journal of Psychology and Theology* 25, no. 2, (1997): 240–253.
>
> Will Slater and Todd Hall, "Measuring Religion and Spirituality: Where Are We and Where Are We Going?" *Journal of Psychology and Theology* 29, no. 1,(2001): 4-21.

Spiritual Well-Being Scale

Abstract. This scale was developed to measure several aspects of the religious dimension of personality. It assists in providing a better understanding of the relationships between religiosity and psychological health. Half of the items in this instrument measure the vertical dimensions of relationship to God; they comprise the Religious Well-Being (RWB) subscale. The other half of the items measure the horizontal dimension of meaning, ideals, faith, commitment, and purpose in life, which form the Existential Well-Being (EWB) subscale. This tool demonstrates preliminary validity, but further validation with conceptually related measures and with known or criterion groups is needed. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. *Journal of Psychology and Theology* 11, no. 4 (1983): 330–340.

Comment. While the MMY reviews indicate that the Spiritual Well Being Scale yields consistent scores and the items are logically related to what it aims to measure, one review questions its use to select spiritual leaders because it does discriminate at the upper end of its score distribution.

> Perhaps because the SWBS [Spiritual Well-Being Scale] is so simple, direct, and easily scored, it is also easily faked. The authors report a negatively skewed distribution in its

scores. This means that SWBS is unable to provide the discrimination typically desired, except to identify those from opposing sectarian or clinically troubled groups. Thus, the SWBS would not be useful in the selection of a spiritual leader from a religious group. Rather, as Bufford, Paloutzian, and Ellison (1991) succinctly indicate, "the scale is currently useful for research and as a global index of lack of well-being" (p. 56). *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

(Psychological) General Well-Being Schedule (GWBS)

Abstract. The (Psychological) General Well-Being Schedule (GWBS) is designed to assess selected aspects of self-representation of well-being and distress. The GWBS can be self- or interviewer-administered and takes eight to fifteen minutes to complete. It includes twenty-two items for six states: anxiety, depressed mood, positive well-being, self-control, general health, and vitality. The six subscales used to measure these states are scored on a Likert-like scale. The GWBS has been used in Great Britain as a measure of outcome of depression. The GWBS is also known as the Psychological General Well-Being Index or Schedule. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. A. F. Fazio, *Vital and Health Statistics, Series 2, Data Evaluation and Methods Research*, no. 73 (Washington DC: U.S. Dept. of Health, Education, and Welfare, Public Health Service., 1977). A Concurrent Validational Study of the NCHS General Well-Being Schedule. Hyattsville, Maryland.

Comment. No MMY review.

Duke Health Profile (DUKE)

Abstract. The Duke Health Profile (DUKE) was designed to measure adult self-reported functional health status quantitatively during a one-week time period. It contains seventeen items rated on a three-point, Likert-type scale. It covers: physical health, mental health, social health, perceived health, and sociability scales. Reliability and validity are discussed. It is available in Chinese, Dutch, French, German, Italian, Polish, Spanish, Swedish, and Norwegian. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. George R. Parkerson, Jr., Box 3886, Duke University Medical Center, Durham, NC 27710.

Comment. No MMY review.

Possession Satisfaction Index (PSI)

Abstract. The Possession Satisfaction Index (PSI) derives from the constructs of materialism and attitude toward money. The PSI is composed of twenty statements scored on a five-point Likert scale from strongly disagree to strongly agree. It measures five factors in possession satisfaction: (1) what possessions can do, (2) what possessions cannot do, (3) public image, (4) success equals possessions, and (5) more is better. Scores on the factors are derived by summing individual item scores within factors. An overall PSI score is obtained by summing scores on all twenty items. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Bearden, William O., et. al., *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research* (Newbury Park, CA: Sage Publications, 1993).

Comment. No MMY review.

Defining Issues Test

Abstract. Measures moral judgment concerning social issues. Multiple-choice test gives information about the process by which people judge what ought to be done in moral dilemmas. Helps understand how people think about social problems. For individuals aged 13-65. Suitable for group use. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Center for the Study of Ethical Development; University of Minnesota, c/o James Rest, 141 Burton Hall, 178 Pillsbury Drive SE, Minneapolis, MN 55455.

Comment. The Defining Issues Test (DIT) is thoroughly reviewed in MMY. It is a reliable survey of values, and beyond logical or face validity, its evidence of validity is differentiation among high school, college, and graduate school students. It is based on Kohlberg's theory and approach to assessing moral reasoning. Items require one to deal with six stories of moral dilemmas. Given the date of its latest edition (1979), today's students may not be familiar with some of its stories. A case in point is the sixth story, "Student Takeover," set in the 1960s. Harvard's Students for a Democratic Society led an occupation of Harvard's administration to protest having ROTC on campus. Nevertheless, the DIT is one of the more widely used measures of ethical and moral reasoning. If a seminary considers it as an assessment of moral judgment, some faculty might review it and even respond to it. Also someone might summarize the MMY reviews.[21]

Rokeach Value Survey

Abstract. The Rokeach Value Survey (RVS) requires one to rank eighteen terminal values, life goals, and eighteen instrumental values, on how one should behave. The eighteen terminal values are: comfortable life, exciting life, sense of accomplishment, world at peace, world of beauty, equality, family security, freedom, happiness, inner harmony, mature love, national security, pleasure, salvation, self-respect, social recognition, true friendship, and wisdom. The instrumental values are those that describe an individual's beliefs concerning desirable modes of conduct—including being ambitious, broadminded, capable, cheerful, clean, courageous, forgiving, helpful, honest, imaginative, independent, intellectual, logical, loving, obedient, polite, responsible, and self-controlled.

Availability. Consulting Psychologists Press Inc., 577 College Avenue, Palo Alto, CA 94036-1490.

Comment. My RVS experience confirms Susan M. Brookhart's observation:

> . . . the RVS is a devilishly difficult task for a respondent with a strong value system. The manual author does note that most respondents reported the ranking task to be difficult; this reviewer thinks that is an understatement.[22]

The MMY reviewers conclude that the RVS's reliability is sufficient to yield stable scores for groups but not for individual assessment. Brookhart concludes her review as advice:

> In sum, the RVS continues to be a useful research tool for theory development about the nature of human values. Its moderate reliability precludes its being used for any purposes that include making decisions with direct consequences for individuals. The RVS should be used with literate adults who are used to dealing with abstractions. The norms are based on data from 1968 and thus are of limited utility.[23]

Artifact Evidence

It seems that some seminaries have extensive untapped resources for assessing outcomes on Personal and Spiritual Formation. These resources are records of periodic, community reviews of MDiv students' intellectual, personal, spiritual, and leadership development and formation. For example, Emmanuel School of Religion, University of St. Mary of the Lake Mundelein Seminary, and St. Vincent de Paul Regional Seminary have systematic procedures to track, discuss, and record students' formation. Robert Hull (Emmanuel's dean), John Canary (Mundelein's rector/president), and Keith Brennan (St. Vincent's rector/president) have graciously shared descriptions of their student development reviews. No doubt many seminaries have reviews along the same lines; I happen to have some knowledge of these three. The purpose in discussing these three is not to suggest that they should serve as models for other seminaries, but the information that such reviews generate could be very useful as outcomes evidence.

Emmanuel School of Religion. Each year in a faculty retreat, the Emmanuel faculty corporately reviews and offers consensus comments on each student's progress. This approach is described in an article by Jack Holland in the ATS journal, *Theological Education* 42, no. 2, 2006. For this Web resource, Robert Hull provided an overview of the Emmanuel approach.

> We're trying to build more rigor and objectivity into a process that has depended too much on anecdote and impression. Briefly, what we do is this: In the spring we take a block of time (usually 10:00 a.m. to 3:00 p.m.) at a site near Emmanuel (often a classroom of a local church) and work through the whole roster of students, with a view to giving a collective evaluation of their progress, as described in the catalog, 28–29 (available on our Web site). A couple of weeks prior, the director of institutional research circulates the roster, with total hours accomplished, current GPA, and space for comments by faculty. After she receives all the comments back, she collates and summarizes these in the last column and sends the document again. Obviously, not all of us will have had substantive contact with every student, but when we gather we make a good-faith effort to generate a collective evaluation. It then falls to me to edit these summary comments into a memo, which is sent, not to the student, but to the advisor. The student is notified by email that the progress review has been completed and instructed to see his or her advisor. A copy is placed in the student's file.[24]

Following the review, each student receives a memo from Hull that reminds the student that the Emmanuel catalog requires the faculty to assess each spring "each student's readiness to continue in pursuit of the stated degree goal." His memo indicates the student's academic standing along with general comments about his or her development. In addition to the notes taken by the student, the seminary maintains a four-column document summarizing academic progress and observations; for spring 2004, seventy-six students were listed. It is formatted as shown below.

STUDENTS ENROLLED IN SPRING 2004
Excluding Students Completing Course Work[25]
MDiv Students

Name	Number of Hours End of Spring Semester	GPA at End of Winter Session	Comments

<u>Mundelein Seminary</u>. The Mundelein Seminary review process has a much longer history than Emmanuel's, and it is deeply rooted in the Roman Catholic theology and practice of priestly formation: spiritual, intellectual, and pastoral.[26] In brief, John F. Canary, rector/president, describes the Mundelein review system as follows:

> Each seminarian is assigned a contact person who usually works with this same student throughout his time in seminary. This allows for a growing trust and accumulated wisdom.
>
> At the beginning of each year, the student writes a series of personal, spiritual, intellectual, and pastoral goals, which must be discussed with the contact and mutually approved, and then these goals are formulated into a formation covenant.
>
> The student meets with the contact twice a quarter (six times a year) to review the covenant and to discuss and evaluate his progress and/or challenge in achieving these goals.
>
> One additional source of input into the formation covenant comes from peer feedback. Each year every student selects six to ten fellow seminarians and requests their feedback as to his progress in preparing well for priesthood. The feedback describes "what is he doing well and where he needs improvement." This honest feedback has been most helpful to the student because it comes from a peer.[27]

Each year, each seminarian's sponsoring bishop and vocational director come to Mundelein to meet with the seminarian and also with the president, dean, student dean, and the seminarian's contact to discuss the seminarian's progress. Before each visit, the contact writes a report based on the seminarian's academic records, comments from teachers, discussion with the *camerata* leader, and his/her general observations. (Every seminarian lives in his own residential hall *camerata* or group composed of about sixteen seminarians and a Mundelein faculty member or administrator.) Over four years, an extensive set of documents accumulate on each student.

<u>St. Vincent de Paul Regional Seminary</u>. St. Vincent has a similar assessment process that is described in more detail below. Assessment occurs in external and internal forums.

External Forum Assessment. The external forum deals with assessment of student formation that determines the student's continuation in the seminary and eventually conferring Holy Orders. These assessments become a substantial part of the student's record and are managed by the student's formation adviser and reviewed by the formation team. The formation team is composed of the vice rector, the deans of formation (one dean for each formation area—human, spiritual, academic, and pastoral formation), and appointed faculty members. Every student has a formation advisor who interviews him monthly and observes his life in the seminary community.

Each fall, the seminarian in consultation with his formation adviser develops specific goals in each formation area: human, spiritual, academic, and pastoral. Goals for the coming year are related to the preceding spring's formal evaluation. Each student's formation adviser presents an informal evaluation of the student to the rector and the formation team. The rector and formation team review the fall informal evaluation along with the preceding spring's formal evaluation, described later, and the student's goals for the coming year. In all student reviews, the rector and formation team may discuss any information derived from the formation adviser's monthly discussions with the student and observations of the student's seminary life.

Each spring, every student is formally evaluated. This evaluation begins with the student writing a self-assessment in which he takes into account his formation adviser's advice and comments from the formation team. The formation adviser then writes his or her formal evaluation, which the student can review before it is sent to the rector and formation team. The rector and formation team comment on this draft. Then the formation adviser writes a final report, which the student is asked to read and sign. If a student disagrees with the final report, he may submit a written report of his disagreements, which is attached to the final report.

Given its discussion of this information, the formation team votes on the student's promotion to the next year and, at the appropriate time, on conferring Holy Orders. Based on all of this, the rector writes an annual report on each student to his respective bishop.

The third year is an in-the-field, pastoral year in which the student develops a learning agreement with his pastoral supervisor. In the fall, the supervisor informally evaluates the student's progress on the learning agreement. Five parishioners at the pastoral site evaluate the student, and the student evaluates himself. All of these evaluations are incorporated in the formation advisor's formal spring evaluation.

Internal Forum Assessment. Any evaluation of spiritual formation drawn from a student's discussions with his spiritual director is held in strict confidence. In most cases, the student's spiritual director is his confessor. Nothing from the internal forum is included in external forum assessment.[28]

Comment. While the Emmanuel, Mundelein, and St. Vincent review processes are intriguing in themselves, their greatest value for evaluation purposes lies in the documents they generate. These documents can be reviewed as artifacts of student learning and formation outcomes because they are gleaned systematically from multiple sources and reflect consensual agreements. Done well, this approximates Donald Schon's *The Reflective Practitioner*[29] done corporately. The documents can be reviewed by juries of faculty and outside readers against an MDiv program's student learning and formation goals. Too often, in my experience, academics adopt the popular assessment techniques of the moment rather than reflecting on substantive material already at hand. Such corporate observation and reflection seems to be the only way to approach assessing the ineffables.

For suggestions on how the documents involved in tracking students may be reviewed as artifacts of learning and formation, see the sections on Assessment Artifacts, Distinguish Exceptional from Typical Outcomes, and Atomistic and Holistic Marking in the ATS Commission on Accrediting's *Handbook of Accreditation*, Section 8, "A Guide for Evaluating Theological Learning."

Personal and Spiritual Formation Assessment Summary. Assessment as we know and experience it is rooted in the Enlightenment, operates rationalistically, and depends on empirical information. Consequently, it is no surprise that conventional assessment strategies and techniques are not at ease with the outcomes of formation as usually understood in religious communities. Nevertheless, we see useful instruments described in *Measures of Religiosity*, the two documents dealing with the relationship of religiosity and spirituality with health—"Scientific Measures of Religiosity and Spirituality," *Multidimensional Measurement of Religiousness/Spirituality for Use in Health Research*, and in the specific instruments briefly reviewed. As described in the section just above, Artifact Evidence, some seminaries already possess an artifact resource from which outcomes evidence can be extracted. It seems to me such resources and what they can yield deserve the time and effort required to discern patterns of outcomes in them.

CAPACITY FOR MINISTERIAL AND PUBLIC LEADERSHIP goal of MDiv Degree Program Standard A

The content area for Ministerial and Public Leadership reads as follows in the standard:

> A.3.1.4 *Capacity for Ministerial and Public Leadership*
> The program shall provide theological reflection on and education for the practice of ministry. These activities should cultivate the capacity for leadership in both ecclesial and public contexts.
> A.3.1.4.1 The program shall provide for courses in the areas of ministry practice and for educational experiences within supervised ministry settings.
> A.3.1.4.2 The program shall ensure a constructive relationship among courses dealing primarily with the practice of ministry and courses dealing primarily with other subjects.
> A.3.1.4.3 The program shall provide opportunities for education through supervised experiences in ministry. These experiences should be of sufficient duration and intensity to provide opportunity to gain expertise in the tasks of ministerial leadership within both the congregation and the broader public context, and to reflect on interrelated theological, cultural, and experiential learning.
> A.3.1.4.4 Qualified persons shall be selected as field supervisors and trained in supervisory methods and the educational expectations of the institution.
> A.3.1.4.5 The institution shall have established procedures for selection, development, evaluation, and termination of supervised ministry settings.

It is with no little trepidation that I attempt to identify instruments for assessing "capacity for ministerial and public leadership." My first concern is the sheer amount of material available on leadership. According to Books In Print, 2,276 leadership books were published in 2005: 28,243 are in-print, and when out-of-print is included with in-print, there are 41,085. A second concern is that valid instruments are built from operational definitions. In this case, one needs operational definitions for "capacity," "ministerial," "public," and especially "leadership." Capacity might be understood as a set of traits that correlate with leadership; put in social science language, a set of independent variables that correlate with the dependent variable of leadership, however defined. To my knowledge, there is no one set of traits that consistently predicts leadership despite the voluminous research on leadership. Beyond not having a set of independent, predictor traits, definitions of leadership vary by values, cultures, institutions, anthropologies, and circumstances. In the context of ministry some may find a consistent view of the leader as servant and shepherd within Scripture and tradition; others may not. And among those who agree with the general shepherd and servant motif, its interpretation and application varies.

With all of the above in mind, several leadership assessment strategies and instruments are identified with *no* attempt to connect them with any particular leadership philosophy or style. As before, the surveys and tests described below were found in the ETS Test Collection; the Abstracts and Availability information are copied directly from the Test Collection. When an instrument is reviewed in *Mental Measures Yearbook*, the MMY reference is cited and observations and comments from it are included.

Campbell Leadership Descriptor

Abstract. The Campbell Leadership Descriptor was designed as a self-assessment of leadership skills as the basis for leadership training. It consists of forty adjectives focused on a specific leadership characteristic grouped into nine major components. Each adjective is given a four-point, Likert-type scale of comparison to self. The areas include: vision, management, empowerment, diplomacy, feedback, and entrepreneurialism. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Jossey Bass, 989 Market Street, San Francisco, CA 94103-1741. 800-274-4434. In Canada call 800-567-4797.

Comment. No MMY review.

Campbell Leadership Index

Abstract. This is an adjective checklist designed to be used in the assessment of leadership characteristics. It has 100 adjectives listed in alphabetical order, each with a definition ranging from "Active—In motion, on the go" to "Witty—Clever and amusing with words." Respondents are asked to indicate on a six-point scale ranging from "Always" to "Never" how descriptive each adjective is of them. Three to five observers also are asked to rate each respondent using the same 100 adjectives and the same scale. The resulting profile shows the comparison on five orientations: leadership, energy, affability, dependability, and resilience. It is a component of the Campbell Work Orientations (CWO), a collection of surveys focusing on the psychological aspects of the working environment. Reliability and validity data are provided. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes: click: Public Access; insert: Title*

Availability. National Computer Systems, Inc., PO Box 1416, Minneapolis, MN 55440.

Comment. Donald Campbell is one of the most eminent psychometric scholars and practitioners; aware of this, one expects much of the CLI. Unlike so many leadership tests, it is reviewed in MMY; these two MMY reviews are recommended to potential users. Both

MMY reviewers found CLI's reliability adequate. George Domino, one of the MMY reviewers, commented as follows:

> Extensive reliability data, including alpha coefficients, interrater reliability coefficients, and test-retest coefficients are presented, generally suggesting adequate to excellent reliability.[30]

On the other hand, its validity evidence is mixed. Domino observes:

> The manual author states that "formulating a simple, general statement about the validity of the CLI has proved to be a difficult task" (p. 129). That indeed seems to be the case, although there is an extensive discussion of various types of validity.[31]

Domino concludes his review with the following sobering comments:
Given Campbell's eminence, it seems unusual the manual cites no published studies.

> Given that the CLI "has been under development for the better part of a decade" (Preface) and that its author is a most prolific and competent psychologist, the lack of references to peer-reviewed papers is particularly disturbing.[32]

As in the case of all self-report assessments, respondents can easily fake it. And is also often true of these instruments, the evidence of validity is weak.

> . . . although considerable validity data are presented, there is no evidence to show the relationship of the CLI to any other test measuring dimensions of leadership, nor to any of the standard personality inventories such as the CPI (California Psychological Inventory), that might be relevant to further establish the construct validity of the CLI. Furthermore, no evidence is presented against real life criteria of leadership either in a concurrent or predictive fashion.[33]

Domino concludes that the CLI compares favorably with similar types of instruments despite its limitations.

> In summary, there are many psychological testing programs on the marketplace that have been designed for use in the business environment. It is my impression that most lack the kind of reliability and validity evidence that can be found with the CLI. Thus, the CLI seems to be a real contribution, though more evidence needs to be presented to satisfy academic and professional users.[34]

<u>Post Heroic Leadership: Managing the Virtual Organization</u>

<u>Abstract</u>. The Post-Heroic Leadership Assessment instrument measures the leadership qualities that are necessary for success in today's changing organizations. It is a self-assessment that contains forty items and uses a six-point Likert-type scale. It has four scales: omnipotence/empowerment, risk taking/rightness, face saving/participation, and co-dependency/development. Suggestions for self-improvement action plans are provided. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: <u>Public Access</u>; insert: <u>Title</u>*

<u>Availability</u>. HRD Press, 22 Amherst Road, Amherst, MA 01002. 800-466-4401.

Comment. Given the current efforts to flatten organizations and my own interest in leadership models that contrast with those described in *The Iliad* and *The Prince*, this test was included because of its title. But I have no experience with it, and MMY has no review of it.

Profile of Leadership Opportunities

Abstract. The Profile of Leadership Opportunities (POLO) was designed to evaluate ten leadership domains to help focus leadership skills where they can make the greatest difference. It consists of sixty items about one's employees rated on a five-point, Likert-type scale. It covers: commitment, communications, leader's personal perspective, supervision, innovation, the workplace, transitions, encouragement, decisions, and employee development. Leadership activities are provided.

Availability. HRD Press, 22 Amherst Road, Amherst, MA 01002-9709. 800-822-2801.

Comment. No MMY review.

Leadership Practices Inventory—Individual Contribution, *Second Edition*

Abstract. The Leadership Practices Inventory—Individual Contribution, Second Edition (LPI-IC), is a thirty-item, ten-point Likert-type rating scale that is used to evaluate individuals' leadership behavior. There is a self-rating scale as well as an inventory for observers (members of an individual's team or those who observe the individual in a leadership role) to complete. LPI-IC coves five leadership practices: challenging the process, inspiring a shared vision, enabling others to act, modeling the way, and encouraging the heart. *ETS TestLink at http://sydneyplus.ets.org/login.asp?timeout=yes : click: Public Access; insert: Title*

Availability. Jossey-Bass Pfeiffer, 350 Sansome Street, 5th Floor, San Francisco, CA 94104.

Comments. The two MMY reviewers are very positive about the LPI-IC. John M. Enger summarizes his comments as follows:

> Overall, Kouzes and Posner have developed a very usable and popular Leadership Practices Inventory that has stood the test of time and continues to hold a prominent place in the market of instruments used primarily for formative evaluation of leaders at various levels of an organization.[35]

L. Carolyn Pearson is also quite positive in her review.

> In conclusion, over the years there has been much empirical research to provide evidence of the reliability and validity of scores from the LPI and LPI-IC. Grounded in a solid conceptual framework, research continues to demonstrate not only that the LPI has good psychometric properties, but also that it demonstrates utility in various settings. The LPI-IC, its instructions, and the facilitator's guide are well written and easy to use. The materials provided in the facilitator's guide are excellent for providing workshop training. I highly recommend the LPI-IC as a useful tool for use with managers and in meeting training needs.[36]

<u>360º Evaluation</u>. The 360º evaluation is a very popular strategy for assessing leadership. A 360º evaluation simply means that one's leadership style is evaluated by one's superior, co-lateral colleagues, direct reports, and oneself. In addition to providing a multiperspective assessment, it also allows the one being assessed to reflect on any discrepancies among the multiple assessments, particularly discrepancies between one's self-assessment and the assessments of others.

The Center for Creative Leadership (CCL) is well known for its leadership research and training programs. CCL also offers high quality 360 leadership assessment instruments. And items in these assessments can be tailored to the needs of different organizations. For further information see: 360 BY DESIGN®—Overview at http://www.ccl.org/leadership/assessments/design360Overview.aspx?SEARCHBTN.X=7\&SEARCHBTN.Y=5&pageId=47

<u>Summary for Capacity for Ministerial and Public Leadership</u>. The list of leadership assessment instruments above represents only a small number of the many that are available. Most that are available reflect the purposes and nature of for-profit or nonprofit organizations with full-time employees. Such instruments are rarely useful for assessing Capacity for Ministerial and Public Leadership. I am certain that there will be suggestions from users of the ATS Web site for additional instruments and different strategies for assessing leadership in ministry.

Endnotes

1. http://www.unl.edu/buros/bimm/html/catalog.html.
2. Ibid.
3. Ibid.
4. Episcopal Church, http://www.episcopalchurch.org/1521_6636_ENG_Print.html.
5 . PC(USA) Presbyteries' Cooperative Committee on Examinations, http://www.pcusa.org/exams/ordination.htm.
6. Ibid.
7. PC(USA) Presbyteries' Cooperative Committee on Examinations, http://www.pcusa.org/exams/bible.htm.
8. Lee S. Shulman, "Making Differences: A Table of Learning," *Change* 34, no. 6 (Washington, DC: Heldref Publications, November/December 2002).
9. Erika Lindemann, *A Rhetoric for Writing Teachers* (New York: Oxford University Press, 2001), 245.
10. Ibid., 245–246.
11. Ibid., 246. Cooper citations: Charles R. Cooper and Lee Odell, eds., *Evaluating Writing: Describing, Measuring, Judging* (Urbana, IL: NCTE, 1977) and Charles R. Cooper and Lee Odell, eds., *Evaluating Writing: The Role of Teacher's Knowledge about Text, Learning, and Culture* (Urbana, IL: NCTE, 1998).
12. Ibid., 246.
13. Professor, Department of Religion and Project Director, Resource Center for Pastoral Excellence, Samford University, Birmingham, Alabama.
14. Penny Long Marler, Professor of Religion, Samford University, Birmingham, Alabama, email correspondence, September 19, 2005.
15. Carl Dudley and Nancy Ammerman, *Congregations in Transition: A Guide for Analyzing, Assessing, and Adapting in Changing Communities* (San Francisco: Jossey-Bass, 2002).
16. Provost and Senior Vice President, Fuller Theological Seminary, Pasadena, California.
17. Sherwood Lingenfelter, mail correspondence, December 7, 2005.
18. Nichole Egbert, Jacqueline Mickley, and Harriet Coeling. "Scientific Measures of Religiosity and Spirituality: Assessing a Missing Component in Health Communication Research," *Health Communication* 16, no.1, 7–27. Copyright © 2004, Lawrence Erlbaum Associates, Inc. Requests for reprints should be sent to Nichole Egbert, School of Communication Studies, PO Box 5190, Kent State University, Kent, OH 44242. Email: negbert@kent.edu Web site: http://www.leaonline.com/doi/abs/10.1207/S15327027HC1601_2;jsessionid=iJWNN8cU08X9e3FjfH?cookieSet=1&journalCode=hc
19. Fetzer Institute, *Multidimensional Measurement of Religiousness/Spirituality for Use in Health Research: A Report of the Fetzer Institute/National Institute on Aging Working Group*. Supported by the Fetzer Institute in Collaboration with the National Institute on Aging (Kalamazoo, MI: Fetzer Institute, 2003 [1999]). http://www.fetzer.org/PDF/Total_Fetzer_Book.pdf.

---

20. Todd Hall, *The Spiritual Transformation Inventory: A Web-Based tool for Spiritual Assessment and Formation* (Concentus Assessment Solutions, Inc., 2005).
21. *Mental Measurements Yearbook 11* http://web5.silverplatter.com.ezproxy.samford.edu/webspirs/start.ws?customer=c45627&databases=S(YB)
22. *Mental Measurements Yearbook* 12 http://web5.silverplatter.com.ezproxy.samford.edu/webspirs/start.ws?customer=c45627&databases=S(YB)
23. Ibid.
24. Robert Hull, Provost, Emmanuel School of Religion, email correspondence, December 20, 2005.
25. Ibid. Attachment.
26. *Program of Priestly Formation*, United States Conference of Catholic Bishops (Washington, DC: 1993).
27. Fr. John F. Canary, Rector of University of St. Mary of the Lake/Mundelein Seminary, Mundelein, IL, email correspondence, December 20, 2005.
28. Fr. Steven Olds, Spiritual Director and Associate Professor, St. Vincent de Paul Regional Seminary, Boynton Beach, FL, letter, January 12, 2006.
29. Donald A. Schon, *The Reflective Practitioner: How Professionals Think in Action* (New York, NY: Basic Books, 1983).
30. Mental Measures Yearbook 12 http://web5.silverplatter.com.ezproxy.samford.edu/webspirs/start.ws?customer=c45627&databases=S(YB
31. Ibid.
32. Ibid.
33. Ibid.
34. Ibid.
35. Mental Measures Yearbook 14 http://web5.silverplatter.com.ezproxy.samford.edu/webspirs/start.ws?customer=c45627&databases=S(YB
36. Ibid.