United States Society and Politics Test (USSP)

Test Manual

Madison Assessment

THE CENTER FOR ASSESSMENT & RESEARCH STUDIES

Revised April 2012

United States Society and Politics (USSP) Version 1

The USSP Test Manual

Section 1.   The Nature of the Instrument

The United States Society and Politics Test (USSP) is a 50-item multiple-choice test developed by the content experts and assessment specialists at James Madison University (JMU). This instrument was designed to assess college students' general education knowledge and goals in the area of American political science and history.

Section 2.   Intended Use

2.1.   Appropriate and inappropriate uses and interpretations

This instrument was designed to assist faculty in direct assessment of student knowledge pertaining to American history and politics. The USSP was developed for use at the programmatic level. That is, the USSP test results can be used for programmatic improvement and refining of the assessment process. Thus, any inferences made about learning or mastery should be made *only* in the aggregate.

The USSP was *not* designed for making decisions about individual students. Currently, psychometric properties are not sufficient to support high-stakes classifications for individuals. This instrument was also not intended as a vehicle for providing individual students with feedback about their knowledge of American society and politics. Institutions may choose to provide their students with individual feedback, but results should *not* be used to make any type of high-stakes classification decisions. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2000), test users are responsible for collecting validity evidence for any uses of the test other than those recommended here. Refer to Section 5.4 *Evidence of reliability* and Section 5.5 *Evidence of validity* for a more thorough discussion of the psychometric properties of the USSP.

2.2.   Target population

The USSP was designed with the focus on undergraduate college students at a traditional four-year institution. However, the USSP is appropriate for the use with any population as long as the learning objectives covered by the test coincide with those set forth by the program. Therefore, in order to determine the appropriateness of the USSP, test users should carefully consider the learning objectives assessed by the USSP. See Table 1 in Section 3.4 T*est design* for the list of learning objectives assessed by the USSP. In addition, test users are encouraged to evaluate the USSP test items for appropriate difficulty level and content. Again, understanding the assessment needs of a program as well as the population of interest is essential for determining the appropriateness of the USSP.

2.3.   Qualifications of users

Proctors administering the USSP should be familiar with administering assessments in standardized conditions. Refer to section 4.1 *Proctor qualifications and training* for more information on proctor qualification. Furthermore, test users should be knowledgeable about how to interpret the statistical results from the test and how to make appropriate inferences about the program using the results. Test users who do not have a measurement background or do not have in-depth knowledge of the program are encouraged to consult with colleagues who have the necessary knowledge.

Section 3.   Test Development

3.1.   Academic and theoretical basis

The USSP was designed to evaluate student learning in five content areas related to identifying, conceptualizing, and evaluating American society and political history. These five areas are listed below:

(1) Social and political processes and structures using quantitative and qualitative data
(2) Key primary sources relating to American history, political institutions and society
(3) The nature and development of the intellectual concepts that structure American political activity
(4) The history and operation of American democratic institutions
(5) The history and development of American society
(6) The history and development of American involvement in world affairs

3.2.   Item type selection

All USSP items are selected response, with the number of response options ranging from four to five. Selected-response items were written to ease scoring, maintain objective scoring, and minimize test-taker fatigue. All items follow a typical multiple-choice format, in which an

item stem is followed by alternative responses consisting of the correct answer and several distracters. An effort was made to create high quality distracters in order to minimize guessing and better differentiate between levels of mastery.

3.3. Item pool and scale development process described

In the summer of 2001, several subject matter experts (faculty specializing in political science and history) convened to craft items of the USSP. Items were written by faculty in direct relation to the selected general education objectives (which are outlined in section 3.1). Sixty of these items were administered during the first pilot of the test in spring 2009. A subset of these items, plus seven items previously used on a similar test, were then selected based on the empirical results and administered during the second pilot. The second pilot test form was later released as the operational form of the USSP.

3.4. Test design

The test blueprint for the USSP appears in Table 1. Some items are mapped to more than one objective; thus, the number of items assessing each objective sums to a value greater than the total number of items on the test.

Table 1

Student Learning Objectives	Test Blueprint for the USSP Pilot (60 items)		Final Test Blueprint for the USSP (50 items)	
	% of Items[a]	Items as numbered on the USSP (60 items)	% of Items[a]	Items as numbered on the USSP (50 items)
1. Social and political processes and structures using quantitative and qualitative data	11.67%	9, 25, 26, 27, 28, 49, 52	16%	5, 9, 24, 25, 26, 27, 42, 44
2. Key primary sources relating to American history, political institutions and society	31.67%	1, 2, 3, 4 ,5 ,6 ,7, 8, 10, 21, 22, 29, 30, 31, 32, 33, 40, 56, 59	22%	1, 6, 7, 8, 10, 21, 22, 28, 29, 35, 47
3. The nature and development of the intellectual concepts that structure American political activity	30.00%	13, 14, 16, 17, 21, 22, 29, 30, 31, 32, 33, 36, 40, 43, 48, 54, 55	24%	16, 17, 19, 21, 22, 24, 28, 29, 35, 41, 45, 46
4. The history and operation of American democratic institutions	48.33%	10, 11, 12, 15, 17, 18, 19, 20, 23, 24, 27, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 51, 52, 53, 57, 60	44%	10, 14, 15, 18, 19, 20, 23, 26, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 43, 44, 48, 50
5. The history and development of American society	16.67%	4, 5, 6, 7, 8, 9, 48, 50, 58, 59	22%	2, 3, 4, 5, 6, 7, 8, 9, 41, 47, 49
6. The history and development of American involvement in world affairs	0%		8%	4, 11, 12, 13

[a] Some items correspond to more than one objective; therefore, the number of items assessing each objective sums to a value greater than the total number of items assessing knowledge of American history and politics.

Section 4.   Administrative Procedures

4.1. Proctor qualifications and training

While administration of the USSP does not require intensive training, proctors should be given guidance on standardized test administration. Proctor training can be accomplished in a brief session in which the proctors are familiarized with the test instructions and the general procedures to be adhered to during the test administration. During training, proctors should be provided with the standardized instructions to be used in the actual testing session. Instructions for each mode of administration are provided in the following section.

4.2.   Testing procedures

The USSP can be administered in either paper-and-pencil or computer-based formats. Although the USSP has not yet been administered as a computer-based test, such mode of administration is available. Section 4.2.2 *Computer-based administration* provides proctor instructions to be used with the computer administration of the test. Issues pertaining to the exchangeability of the two administration modes are discussed in section 4.3. Room temperature and lighting should be appropriate for optimal performance testing. Before beginning the test, examinees should be provided with general information about the number, type, and content of items on the test. Examinees should be informed of the amount of time they will be given to complete the test and what they should do upon completion of the test. It is recommended that students be given at least 40 minutes to complete the USSP. However, if the majority of students are still working after 40 minutes, the proctor can extend the testing period for another 5 minutes. The proctor should periodically announce the time remaining. After 30 minutes have passed, inform the students that they have 10 minutes remaining. After 35 minutes have passed, inform the students that they have 5 minutes remaining. Proctors should circulate in the room to remedy problems, answer questions, replace pencils, and generally motivate students to perform their best. Recommended standardized test instructions are provided below. Note that the instructions will change slightly based on the administration mode selected.

The USSP Instructions

> This test is designed to assess your knowledge of United States society and politics. You will have 40 minutes to complete the 50 multiple-choice items on this test. For each question, select the answer that you deem to be the most accurate. Respond to all questions to the best of your ability. I will let you know when you have 10 and 5 minutes remaining. When you are finished with the test, you may review your answers. If you finish early, please be respectful of others and wait quietly for further instructions from the proctor. Please raise your hand if you have any questions. You may begin.

The instructions above were used in the testing sessions leading to the data generated for use in this manual. *If you elect to give your test with different instructions than those presented here, please understand that you will be administering the test under non-standardized conditions. As a result, the scores you obtain may be incomparable to the sample data offered in this manual.*

4.2.1.   Computer-based administration

Students will each be seated in front of a computer on which the USSP test has been installed (*clients will be sent a URL that is specific to that institution)*. Students should receive either oral or written instructions on how to enter their identification information and begin the test. The instructions should also be provided on the computer screen just prior to administering the first item.   This URL is compatible with all browsers. All pop up blockers should be disabled.

4.2.2.   Paper and pencil administration

If paper and pencil administration of the USSP test is elected, additional fees apply. Examinees receive an item booklet and an answer sheet such as a scantron form. If a scantron form is used as an answer sheet, a form that can accommodate up to five response options is required. For example, the General Purpose Form (number 6703) by Pearson NCS has been used successfully for the USSP. Examinees should be given instructions on how to identify themselves for the test (if necessary) and how to record answers on the answer sheet properly. These additional instructions may change depending on the format of the answer sheet used and identifying information needed.

Additional Instructions for Paper-and-Pencil Administration

> First, write in your student id and last name. Make sure to bubble in the letters of your last name and the digits of your id number. For each item of the test, choose the best answer and fill-in the corresponding bubble on the answer sheet with a #2 pencil. Please do not write on the test booklets; answer all questions on the answer sheets provided.

4.3.   Extent of exchangeability

When an instrument is administered in both paper-and-pencil and computer-based formats, information derived from one context may not be directly applicable to another context. In other words, it cannot be assumed that reliability or validity information collected through one mode of administration will generalize to the other mode (Mead & Drasgow, 1993; Wise & Plake, 1990). Equivalence should be established before applying information collected in one setting to information collected in another. Please note that the USSP has only been administered in a paper-and-pencil form to date. Due to the nature of the test items, their transfer to a computer-based platform should be easily achieved. However, future data collected in both formats will provide evidence as to the exchangeability of paper-and-pencil and computer-based formats. Again, test users are encouraged to conduct these exchangeability studies before proceeding to use the test interchangeably in both formats.

Section 5. Technical Information

5.1. Scoring and interpretation

All USSP items are selected response. The majority of items have five response options including the correct response. The range is between four and five response options. Items are scored dichotomously: a correct response to an item is given a score of '1' and an incorrect response to an item is given a score of '0'. The total score is obtained by summing the scored item responses. Higher total scores indicate that examinees have higher knowledge of American political history, and lower total scores indicate that examinees have lesser knowledge of American politics and history. The highest possible score is 50; the lowest possible score is 0. The following section outlines the demographic characteristics of the sample to which the USSP was administered.

5.2. Student sample

The pilot version of the USSP was administered in Spring 2009 and the final version was administered in Spring 2011 and Spring 2012 to random samples of mid-career students at JMU. The total sample size was 685 for Pilot 1 and 1389 for the final version. The ethnic backgrounds of the students in the sample roughly approximated those of the overall university population, with the majority of students being non-Hispanic White. Table 2 presents the racial breakdown of the student sample. Table 3 presents aggregate information about students' gender, age, and GPA.

Table 2
Demographic Information for Student Sample

Race	Pilot	Final Form
Non-Hispanic White	83.0%	77.6%
Black	4.0%	3.4%
Hispanic	2.5%	3.7%
Asian	5.0%	5.8%
Not Specified	5.0%	8.2%
Other		1.3%

Table 3
Gender, Age, and GPA of the Student Sample

Variable	Pilot Frequency (%)	Final Form Frequency (%)
Gender		
F	432 (63.6%)	843 (61.0%)
M	247 (36.4%)	540 (39.0%)
	Pilot Mean (SD)	**Final Form Mean (SD)**
GPA	3.0 (0.6)	3.0 (0.7)
Age	20.2 (1.0)	20.2 (1.3)

Note. N = 679 on the pilot test 1383 on the final form of the USSP because 6 students on each test form could not be linked to their demographic information.

5.2.1. Testing Conditions

The USSP was pilot-tested during Assessment Day, which is a semi-annual university-wide testing session dedicated to obtaining direct evidence of student learning and development. Students take the same tests at the beginning of their academic career as incoming freshmen and then again at the mid-point of their career, when they have earned between 45-70 academic credits. This design allows for capturing growth in student learning and development. Moreover, it enables administrators to evaluate the effectiveness of general education, because most students with 45-70 academic credits are sophomores and juniors who have completed some of the general education requirements. Notably, Assessment Day is a low-stakes testing event. Under the low-stakes testing conditions, students may not give their

best effort; therefore, the scores may underestimate student knowledge of the intended construct. Nonetheless, these data provide valuable insight into the functionality of the USSP.

The next section provides an overview of how this research design enables investigation of the USSP scores.

5.3. Research Design

Assessing students at the mid-point in their academic career allows exploration of the test scores of those who have completed their general education requirements. More specifically, the general education requirements in the area of American Experience are of particular interest in this situation. In order to meet this requirement, students have a choice of completing *one* of the *three* courses covering different aspects of the American life. Table 4 presents these courses and their short descriptions. Since the USSP was designed to assess knowledge of American history and politics, we would expect students who had not completed the American Experience requirement at all to score lower on the USSP than students who had completed this requirement.

Table 4
American Experience Courses and Descriptions

Course	Short Description*	Long Description**
GHIST 225 U.S. History	Contextual and document-based study of the American historical experience that emphasizes the interaction of people, ideas and social movements.	A survey of U.S. history from the Colonial period to the present, emphasizing the development of American civic life, the involvement of the U.S. in world affairs and the cultural richness of the American people. This course stresses the analysis and interpretation of primary sources.
GPOSC 225 U.S. Government	Evolution and contemporary operation of the American political system by examining its fundamental principles and current dynamics.	An examination of institutions, processes and intellectual concepts which structure American political activity. The interaction of the political system with the changing American society and America's changing role in world affairs are also treated. The course provides an introduction to quantitative methodology.
GJUST 225 Justice and American Society	Historic and contemporary events in terms of issues of justice, highlighting how societal structures interact with individual lives and vice versa.	This course introduces the student to the concept and reality of justice in America. It is a broad-based, interdisciplinary consideration of justice: What it is, what it means, and how it intersects with society and social institutions in America. Philosophical and theoretical underpinnings of the notion of justice and the historical context of justice in American society will be considered.

* Short descriptions are from the JMU website
** Long descriptions are from the JMU course catalogue

Alternatively, students can forego enrolling in these courses by satisfying the American Experience requirement through (a) transferring academic credit from a different institution or (b) AP credit from high school. Therefore, the pathways through which students can fulfill the American Experience requirement differ in ways that help us understand what their scores on the USSP mean. For example, it is of interest to explore whether students who transferred credits perform differently on the USSP than students who completed these courses at the home institution. Exploring these kinds of differences and relationships between the USSP test and course grades is one way to garner construct validity evidence for the test. Section 5.5 provides a more elaborate treatment of the findings pertaining to the validity evidence for the test. The following sections discuss evidence garnered for the reliability and validity of the USSP.

5.4. Evidence of Reliability

Reliability refers to the degree of stability and consistency of test scores, as well as the degree to which test scores are free from random error. Due to the various sources of variability in test scores, there are different ways of measuring reliability. The USSP has been examined for internal-consistency reliability as measured by coefficient alpha ( ), which is frequently used to determine reliability.

Specifically, requires only one administration and is the mathematical equivalent of the average of all possible split-half reliability computations. Alpha indicates how much variance in the observed scores is attributable to the true score. In other words, indicates how related the scores on the items are to the primary construct measured by the test. While coefficients with a value of .70 or higher have traditionally been considered adequate for program evaluation or research, reliabilities above .80 are desirable for evaluation of individual students (Nunally, 1978).

The reliability of the USSP test has been calculated based on the samples described in section 5.2. Table 5 presents the observed reliability estimate, along with the descriptive statistics for this sample. Cronbach's alpha ( ) for the USSP was approximately .87, indicating a high level of internal consistency.

Table 5
Coefficient Alpha and Descriptive Statistics for the USSP

	N		**Mean**	**SD**
Pilot 1 (60 items)	685	.85	33.6	9.1
Final Form (50 items)	1389	.87	28.5	8.7

5.5.   Evidence of Validity

Validity refers to the degree to which one can make the desired inferences from the scores obtained on a test. Validity is not an absolute state, but rather a collection of evidence indicating that the scores obtained on a test are valid for their intended use (AERA, 2000). For the USSP, validity evidence was garnered in the following ways. Content validity evidence was ascertained via examination of the item development process, including the objectives targeted by the test and the relevant expertise of item developers. For construct validity, various methods or pathways of identifying students who have accomplished the objectives targeted by the USSP were employed. More specifically, the performance of students on the USSP who have satisfied the American Experience requirement in different ways was examined. Additionally, the correlation between course grades and test performance was computed, to show evidence that the test was measuring concepts related to the coursework. Refer to Section 5.3 for the description of American Experience requirement. The next sections discuss evidence pertaining to content and construct validity of the USSP.

5.5.1.   Content Validity

Preliminary evidence for content validity of USSP is integral to the scale development process. First, the test items were specifically created to assess the learning objectives outlined in Table 1, Section 3.4. Second, all items were written by the subject matter experts teaching general education courses in American politics and history. Therefore, the items of the test can be presumed to accurately reflect the domain of American politics and history.

5.5.2.   Construct Validity

Construct validity evidence was gathered through investigating student's performance on the USSP in relation to students' performance in courses addressing the same or similar objectives as the USSP. Generally speaking, students who completed a relevant general education course satisfying the American Experience requirement should have a *greater* amount of knowledge in this area when compared to students who have not completed the requirement. If a test is found to reflect this difference, then its construct validity is strengthened. Moreover, it is of interest to compare and contrast the USSP scores of students with transfer credits as well as students with AP credits. The next sections discuss the findings pertaining to construct validity of the USSP.

Tables 6a and 6b present average USSP scores for different groups of students: those who had not completed the American Experience requirement, those who had completed it at JMU, and those who had completed this requirement elsewhere.

The USSP scores of students who had not completed the requirement present the baseline information of students' knowledge of American history and politics prior to completing any coursework on this subject matter. The USSP scores of students who had satisfied the requirement either through completing courses at JMU, through AP credit, or by Transfer credit should reflect a gain compared to the baseline USSP score. Therefore, we expect the USSP scores of those students who had completed the requirement would be higher than of those students who had not. Looking at the values presented in Table 6, it is evident the USSP scores of students who had completed the requirement through the AP credit are the highest, followed by the scores of students who finished the relevant course, followed by students currently enrolled in the course (about 6 weeks into the semester). Students with transfer credit scored similarly to students who have not taken any coursework in this area; this matches the findings from a similar test used at JMU to assess the American Experience course.

Additionally, two samples of incoming first-year students took the final form of the test the week before classes started in Fall 2010 and Fall 2011. Their scores are reported in Table 6c. Estimated score reliability was .86, similar to the estimate in Table 5. First-year students with no AP or transfer credit scored similarly to the mid-career students with no course credit. Transfer students scored a bit higher than mid-career students with transfer credit; given the small number of students with transfer credit this could be random fluctuation or it could be that transfer students forget more content over time.

Madison Assessment The United States Society and Politics Test, Version 1 Test Manual- 2011

Table 6a
Pilot Test Scores (preliminary test form) by Student Groups. Maximum score = 60.

Student Groups	N	Mean	SD
Requirement not completed	247	30.43	8.18
Not currently enrolled	146	29.36	8.26
Enrolled in one of the required courses	101	31.97	7.86
Requirement completed at JMU	301	34.23	8.31
Requirement met by AP credit	80	43.55	8.65
Requirement met by Transfer credit	57	30.33	7.09
Total	685	33.62	9.13

Table 6b
USSP Scores (final test form) by Student Groups. Maximum score = 50.

Student Groups	N	Mean	SD
Requirement not completed	489	25.73	7.89
Not currently enrolled	284	25.50	8.27
Enrolled in one of the required courses	205	26.05	7.33
Requirement completed at JMU	587	28.35	8.13
Requirement met by AP credit	190	37.78	6.54
Requirement met by Transfer credit	123	25.79	7.99
Total	1389	28.49	8.74

note: On the pilot version of the test, one student had both AP and transfer credit; this record was coded as AP. Three students had both JMU and transfer credit; their records were coded as completed at JMU. Eight students had both AP and JMU credit; their records were counted as met by JMU credit.
On the final form of the USSP, three students had both JMU and transfer credit; these records were coded as completed at JMU. Ten students had both AP and JMU credit; their records were also counted as met by JMU credit.

Table 6d
Incoming Students USSP Scores. Maximum score = 50.

Student Groups	N	Mean	SD
Requirement not completed	1891	25.78	7.96
Requirement met by AP credit	251	37.86	5.13
Requirement met by Transfer credit	123	27.48	7.57

Another piece of validity evidence is the correlation between course grades and test performance. Students who earned a high course grade should score higher on the test than those who earned a low course grade, if the test is measuring something related to the course. Table 7 presents correlations between grades in two courses, either one of which could be used to satisfy the requirement, and the USSP score. The third course GJUST225 was not included due to insufficient sample size, which would lead to an unstable estimate. Correlations close to 0 indicate no relationship, while correlations closer to 1 indicate a very high relationship between test scores and course grades. Assuming

that these courses and the USSP address the same domain of American history and politics, we would expect the relationships between course grades and the USSP scores to be moderately high.

Table 7
Correlations between Course Grades and USSP Scores

Course	N	Correlation
Pilot: GHIST 225	171	.31*
Pilot: GPOSC 225	145	.49*
USSP (Final form): GHIST 225	349	.14*
USSP (Final form): GPOSC 225	234	.42*

* = p-value < 0.01

As depicted in Table 7, the relationships between both courses and the USSP scores are moderate, suggesting that they both tap into the same construct. The correlation between the average grade in GPOSC 225 and the USSP score appears to be stronger than the correlation between the average grade GHIST 225 and the USSP score. Additionally, the mean score for students with political science credit was about 0.18 standard deviation units higher than the mean score for students with history credit. This suggests that the construct measured by the test may be a bit more related to US political science than to US history at JMU. However, in the graph below, notice the overlaps in the confidence intervals for the correlations.
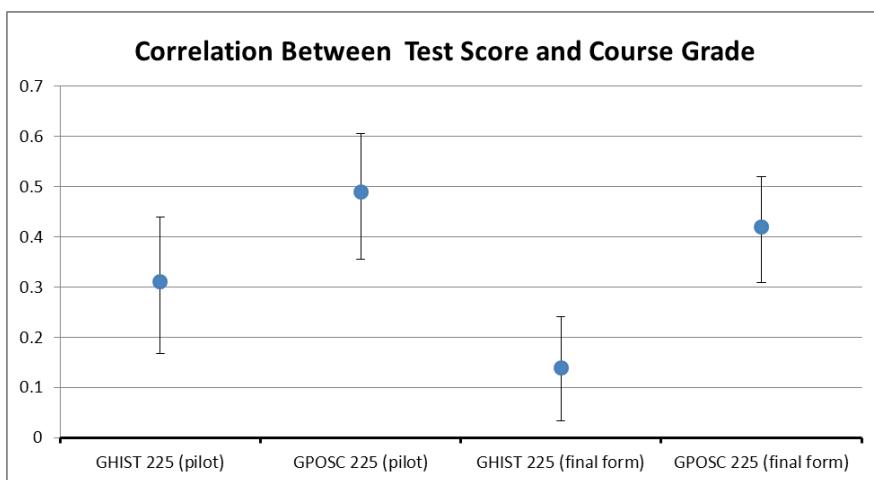

Figure 1.

5.6.   Summary of the validity evidence

In summary, the current analyses provided some evidence for the validity of the USSP. The test scores appeared to differentiate well between students who had completed courses in American history and politics from those who had not. The scores differed depending on the specific route taken by students to satisfy this General Education requirement. For example, students with the AP credits tended to score much higher on the USSP than all other students, whereas students with Transfer credits scored no differently than students at the baseline level. It is unclear whether these findings are indicative of the test functionality, or if they reflect the differences among student groups. For example, the AP students had higher SAT scores as well. More evidence needs to be collected for any conclusive inferences. Further, the investigation of the relationships between General Education course grades and the USSP scores revealed that these relationships were both positive. This finding provided support for the validity of the USSP.

Section 6.   Additional Information

6.1.   Where to get additional information

Additional information on the USSP may be obtained by contacting Madison Assessment (http://www.madisonassessment.com).

Section 7. References

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2000). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Gulliksen, H. (1987). *Theory of mental tests.* Hillsdale, NJ: LEA. (Original work published in 1950)

Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized tests and paper-and-pencil cognitive ability tests: A meta-analysis. Psychological Bulletin, 114, 449-458.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Wise, S., & Plake, B. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development*, *23*(1), 3-10. Retrieved from PsycINFO database.